THE QUARTERLY JOURNAL OF
EXPERIMENTAL PSYCHOLOGY

Vol. II 1950 Part 1

A COMPARISON OF THE INFLUENCE OF HANDLE LOAD AND OF UNFAVOURABLE ATMOSPHERIC CONDITIONS ON A TRACKING TASK

BY

A. CARPENTER

(*From the Medical Research Council Applied Psychology Unit, Cambridge*)

An experiment was made to determine the extent to which the performance of a tracking task was adversely influenced by simulated tropical environmental conditions, by comparing this effect with that of loading the control handle with weights. The pursuitmeter used was similar to that used by Mackworth (1948) when he showed that the accuracy of performance was reduced under tropical conditions. The score of error was the total angular separation between the target and follower pointers expressed in arbitrary units. It was found that as the room temperature was raised from approximately 80° to 90° on the Effective Temperature scale, this score increased in the ratio 1·13, an increase which was also obtained by increasing the load on the control handle from 8 to 30 lb. There were also indications that the effect of room temperature was greater with higher handle loads. The effect on performance was not correlated with rectal temperature or with the amount of weight lost.

I

Introduction

Previous experiments in the Cambridge hot-room, and elsewhere, have shown that in a variety of different tasks, performance becomes less good as the temperature of the room is raised above about 85°. The tasks in which this deterioration has been found have included visual and auditory vigilance (Mackworth, 1950), rapid repetitive highly-learned tasks such as the reception of telegraphy (Mackworth, 1946), and the setting and decoding of messages on a mock-up coding board, purely physical work, using an ergograph, mental work, involved in the tracing of a hidden electrical circuit (unpublished), and a tracking task such as was used in the following experiment.

In all these experiments deterioration in performance at higher temperatures has been found, to an extent which was statistically reliable, and in some cases an approximate assessment of the degree of impairment was possible. For example, in the telegraphy experiment, the errors increased from an average of 3·2 per 1,000 characters at 79° on the Effective Temperature scale, to 25·3 at 97°. The difficulty in attempting to measure the performance on synthetic tasks such as those provided by an ergograph, or the tracking task provided by the pursuitmeter in the following experiment lies mainly in understanding the measuring scale. The proportion of errors made in telegraphy reception is a meaningful measure of performance, but in such a highly artificial task as tracking on a pursuitmeter, the score obtainable on our machines—merely the total angular displacement between the target pointer

and the subject's pointer—is really only of value as a means of comparing one condition with another.

This present experiment is an attempt to give a little more meaning to the extent of deterioration of this tracking task, by comparing the effect of room temperature with that of loading the handle with known weights.

A second purpose in the experiment was to find out whether there was an appreciable difference in the effect of temperature when the load on the handle was varied, that is to say, to discover whether the effect of high temperatures was mainly on physical work, or on the more skilled side of tracking.

II

Apparatus

The pursuitmeter used in this experiment consisted of a display of two pointers, of white ivorine, about 3/16 in. × ¼ in., with a bold central indicator line marked in black ink. These moved along an arc of radius about 8 in., and up to 3 in. in length, the upper being moved irregularly to and fro by an irregular cam operated by an electric gramophone motor, and the lower being operated by the subject by a horizontal handle 2 ft. in length, which was moved up and down through a maximum distance of 18 in. On this handle lead weights could be hung, giving loads, as measured at the end held by the subjects, of 2, 8, 16, 24, 32 and 40 lb. The task was to follow the irregularly moving upper pointer with the lower, keeping the two in alignment.

III

Procedure

Three such machines were used, and three subjects tested at a time. There were four groups of subjects, and each performed the test four times, once at each of four room temperatures, which were 85°/75°, 90°/80°, 95°/85°, 100°/90° (° F. dry bulb/° F. wet bulb), the air velocity being approximately 100 ft. per minute. These four conditions were arranged with respect to subject group and to serial order (called "days") in a Latin Square arrangement.

In any one day each subject did eighteen test spells, each of 3 minutes, which consisted of one spell on each machine with each of the six weights. Within this framework, conditions were randomized.

IV

Subjects

The twelve subjects were physically fit young Naval ratings, aged between 20 and 23 years. They were "acclimatized" to hot-room conditions by spending 3 hours per day for the fortnight before the experiment in the hot-room, at a temperature of 100° F. dry bulb/90° F. wet bulb. They wore shorts only, while in the hot-room.

Rectal temperatures were taken (*a*) on entering the hot-room, (*b*) after 1 hour of acclimatization and (*c*) on completion of the experiment. Fluid lost as sweat during each experiment was assessed by keeping a balance sheet of body weight to the nearest 5 gm.

V

Scoring

The total angular separation, irrespective of sign, between the two pointers of each machine, which occurred during each 3-minute experimental spell, was recorded on a ratchet disc and taken as the error score for that spell.

VI

Results

The experimental design enabled the five major sources of variation in the results obtained, to be separately evaluated; namely, that due to the use of different machines, that between subjects, the effect of practice, of room temperature conditions, and of handle load. Of these, only the last two are experimental variables of interest, and the purpose of separately evaluating the effects of the first three is to exclude the variance of these effects, and to prevent it from reducing the accuracy of the estimation of the effects of room temperature and of handle load. They will not, therefore, be given here.

Table I gives the total error scores obtained at each of the six degrees of handle load and of the four room temperature conditions.

TABLE I

Error Scores

Room Temperature	*Handle Loading in lb.*						*Totals*
	2	8	16	24	32	40	
85/75	3,363	3,259	3,278	3,403	3,614	3,763	20,680
90/80	3,388	3,314	3,362	3,492	3,734	3,862	21,152
95/85	3,762	3,565	3,703	3,989	4,329	4,625	23,973
100/90	3,644	3,588	3,680	3,938	4,030	4,436	23,316
Totals	14,157	13,726	14,023	14,822	15,707	16,686	

(*a*) *Effect of Room Temperature.*

The temperature totals in Table I show a marked increase of error score between room temperatures of 90°/80° and 95°/85°. The standard error of these total figures, taken from an analysis of variance, is 350, and thus this increase has a critical ratio of 8·06 and is highly significant, although the differences between the two lower temperatures and that between the two higher temperatures is not significant.

This agrees very well with Mackworth's figure of 94/84 (° F. dry bulb/wet bulb) as the point at which deterioration in performance with temperature becomes statistically reliable.

An approximate quantitative estimate of the degree of this deterioration can be made by taking the difference between the mean scores at the two lower and the two higher temperature conditions. This gives an increase of error score, for an increase of 10° F. in temperature, in the ratio 1·13.

(*b*) *Effect of Handle Load.*

The variance ratio of the error totals due to the several handle loads is 23·05, and the standard error of any one total is 238. Thus the general trend of increase in error with load is significant, although the drop in error between loads of 2 lb. and of 8 lb. is not so. This drop, however, is interesting in view of Hick's finding (Hick, 1946) that moderate frictional load, in a tracking experiment with crank handles, improved the accuracy of performance, presumably by providing an influence stabilising speed. In the present case a medium weight of 8 lb. might improve performance by reducing rapid small correcting movements, which while not con-

tributing much to a truly integrated score, would be counted fully on a ratchet device which takes no account of duration of error.

If the ratio 1·13 found between the scores obtained at the higher and lower temperatures is compared with the effects of handle load, it will be seen that this same ratio is obtained between the scores at loads of 8 and 30 lb., or between 16 and 32 lb. Thus the effect on performance of increasing the room temperature from 87/77 to 97/87, i.e. by ten degrees, is approximately the same as that of adding a further 20 lb. weight to the load on the pursuitmeter handle.

(*c*) *Interaction of Load and Room Temperature.*

This experiment was designed to provide some evidence as to whether the effect of room temperature was greater with the heavier handles, or, to put it another way, whether the effect of handle load was greater at the higher temperatures.

The figures in Table I show that this is the case, but the effect is rather hidden by the presence of erratic variation. In fact, the analysis of variance showed that the effect we are looking for, the weight versus temperature interaction, had a variance ratio of only 1·44, as compared with the weight versus temperature versus machine interaction, which is non-significant, representing a P of between 0·15 and 0·2. The effect can nevertheless be clearly shown by taking the ratio of the scores at the two higher temperatures to those at the lower two, at each value of weight separately. This gives the results shown in Table II; which show that the increase in error score

TABLE II

Handle load, lb.	2	8	16	24	32	40
Score ratio	1·10	1·09	1·11	1·15	1·14	1·19

brought about by an increase in room temperature from about 80° on the Effective Temperature scale to 90° Eff. is almost doubled by increasing the handle weight from 2 to 40 lb. It is interesting to compare this with Mackworth's results (Mackworth, 1950). Using the same apparatus he found that the effect of temperature was greater with 50 lb. loaded upon the control handle than it was with the heavy handle removed, and the apparatus controlled by a light knob operated by one hand only.

(*d*) *Measurements of Body Temperatures and Weight Loss.*

These measurements were taken as part of a longer-term programme whose aim is, if possible, to correlate change in performance with the physiological effects of heat. As was to be expected, and as we have found in previous experiments in the Cambridge hot-room, both body temperature and weight loss showed, when averaged for all the subjects, a close relationship to room temperature.

TABLE III

	Room Temperatures			
	85/75	90/80	95/85	100/90
Rise in body temperature, ° F. ..	0·45	0·55	0·9	1·2
Loss of weight, gm.	350	500	800	1,100

The relationship between temperature and weight loss and individual performance scores is, however, slight. Product moment correlations of the physiological measures, with the extent to which each individual's score on any one day differed from his own average, gave figures of $r = 0{\cdot}35$, with rise in rectal temperature, and $r = 0{\cdot}15$ with weight loss. Of these, only the first is statistically significant, and that only just so. The relationship between the individual physiological measures and actual performance score, uncorrected in the above way, was still less. It is likely that, such as it is, this correlation is entirely due to the common factor of room temperature, and that there is no direct relationship between performance and either body temperature or rate of sweating. Although the figures from this experiment are far from sufficient in themselves to decide this point, such repeated negative findings suggest that the deterioration in performance is related to room temperature via some mechanism which does not involve a change in such physiological conditions as body temperature or rate of sweating.

VII

The Subjects' Opinions

Between 2 and 5 days after the experiment ended, ten of the twelve subjects were asked to complete a simple questionnaire form, asking for their opinions about the temperature and weight conditions. Comparing these opinions with the results obtained, it is interesting that:

(*a*) eight subjects preferred some weight on the handle to none;

(*b*) the largest weight which it was thought could be controlled without increase in error was between 24 and 32 lb., whereas, in fact, 24 lb. produces an increase in error;

(*c*) eight subjects thought 90/80 was the highest room temperature they could tolerate without increase in error, the other two considered 95/85 was tolerable without decrease in accuracy. The majority vote in this case agrees with the numerical results very well.

VIII

Conclusion

The results of this experiment are in themselves insufficient to show the interaction of handle weight and room temperature to be statistically reliable, and thus they do not add any certainty to the inherent probability of such an effect. In providing a comparison between the effects of handle load and room temperature, however, the result of this experiment was satisfactory.

The most interesting item of information gleaned from the questionnaire replies is the confirmation that the addition of a small weight—about 10 lb.—to a control handle of this type, does lead to some improvement in accuracy of tracking.

IX

Summary of Results

(1) Earlier findings, that performance of the tracking task provided by the pursuitmeter was adversely affected by high environmental temperatures, were confirmed.

(2) The magnitude of the effects of heat and of load are of the same order, and, in fact, the deterioration in accuracy between the two lowest and the two highest temperatures is equal to that between weight loads of 8 and 30 lb., i.e. an increase

of effective temperature from 80° to 90° is equivalent to adding 20 lb. load to the handle.

(3) The results indicate that the effect of loading the handle is greater at higher room temperatures, but this effect is not statistically reliable in this experiment.

(4) Rectal temperature and amount of weight lost in sweat show the expected relationship with room temperature, but the individual readings show no reliable correlation with performance.

X

References

1. Mackworth, N. H. (1950). "Researches on the measurement of human performance." Medical Research Council, Special Report No. 268.

2. Mackworth, N. H. (1946). Effects of heat and high humidity on wireless telegraphy operators' hearing and recording Morse messages. *Brit. J. Indust. Med.*, **3**, 143.

3. Hick, W. E. (1946). "The effect of heavy loads on handwheel tracking." Unpublished report, RNP46/313, to the Royal Naval Personnel Research Committee of the Medical Research Council.

(*Manuscript received 17th June*, 1949.)

AMNESIA AND THE GENERIC IMAGE*

BY

O. L. ZANGWILL

(*From the Institute of Experimental Psychology, Oxford*)

It is pointed out that the fate of generic images has been studied little, if at all, in amnesias of organic origin. Observations are described in a case of alcoholic Korsakoff psychosis which appear to indicate that generic images are subject to Ribot's Law of Regression. The patient, a spinster of 57, with some talent for drawing, was asked to represent a number of objects, e.g. a woman dressed in the latest fashions, which might reasonably be regarded as "generic" in nature. These drawings suggested that the patient's generic images were, in some cases at least, appropriate to an earlier period of her personal history. In general, this ante-dating effect was in good correspondence with the scope of the retrograde amnesia. Some implications of these findings for the psychological theory of memory are briefly discussed.

I

Introduction

The origin and nature of generic images were much discussed in the earlier systematic texts. "An image," wrote Stout, "is called generic when it possesses a distinct and salient centre or core corresponding to the common characters of a class, together with a vague and inconstant margin" (1896, I, p. 179). In particular, the possible role of the generic image in the evolution of the concept aroused lively controversy. This problem, which was admirably discussed by William James (1890, II, pp. 45–50), need not concern us here, and it is sufficient to recall Ward's comment that "The 'generality' of the generic image differs from the true universality of the concept in that the former is the passive and accidental result of reduplication, the latter the product of definite and active comparison" (1918, p. 199). As regards the origin of the generic image in individual experience, it has commonly been held to arise from the blending or fusion of a large number of particular images of identical or related content. As is well known, this view gained considerable prestige as a result of Galton's studies of composite portraiture (1907, pp. 229–33). It is highly doubtful, however, whether the parallel between the generic image and the composite photograph is as close as Galton supposed. In the first place, more recent studies (e.g. those of Katz, 1948) suggest that composite portraits are apt to acquire characteristics lacking in the typical generic image. And in the second place, it is now customary to regard recollection as displaying a more active character than is implied in Galton's treatment.† At the same time few would disagree with Galton's dictum that the character of our generic (or "cumulative") images depends, to a considerable extent at least, upon our nurture (1907, p. 132).

The changes undergone by generic images in the course of life would repay closer study. It is evident, for instance, that the generic images which we entertain in childhood differ appreciably in content from those which we entertain in adult life. Further, our general image of a given individual (or class of individuals) undoubtedly changes in content as we ourselves grow older. If, further, it is agreed that the images in question depend on some kind of cumulative disposition, it might be

* Based on a paper read to the Experimental Psychology Group on 20th July, 1947.

† "Wherever a single cause throws different groups of brain elements simultaneously into excitement, the result must be a blended memory."—Galton (1907, p. 229.) This was written in 1879.

expected that they would show regressive changes in cases of organic amnesia associated with brain injury or disease. The purpose of the present paper is to report some observations in a case of Korsakoff's psychosis which appear to confirm this expectation.

It has long been recognized that organic amnesia displays a retrograde (or retrogressive) tendency. This was formulated by Ribot under the title of the *law of regression* (1885, pp. 121–2). In cases of diffuse brain disease, recent events are the first to be forgotten and, as the disease advances, the amnesia comes to embrace events lying further and further back in the patient's psychological history. In severe cases of Korsakoff's psychosis, it is usual to find a retrograde amnesia whose scope embraces a very considerable proportion of the patient's life history. In the examination of such cases, however, it has been customary to restrict inquiry to the patient's memory for events in his personal life. To the knowledge of the present writer, no investigator has attempted to study the fate of generic images in patients who have sustained prolonged retrograde amnesias. It is felt, therefore, that the present observations, slight as they are, warrant being placed on record.

II

Observations in a Case of Korsakoff's Psychosis

The observations now to be described were made in connection with a more extensive investigation of amnesic syndromes undertaken in 1939–40 and interrupted by the war. The work was carried out at St. Ebba's Hospital, Epsom,* and at the Cambridge Psychological Laboratory. Some reference to the present case has already been made by the writer (Zangwill, 1941) in a paper on paramnesic phenomena in the Korsakoff state.

The patient was a spinster, aged 57. On admission (June, 1939) she was confused, disoriented and grossly amnesic. She made frequent mistakes of identity and confabulated freely. She was in poor general health, with peripheral neuritis in both legs. The mental condition remained stationary for some six months, after which time a certain improvement was manifested. Orientation gradually returned and confabulations became rarer and less spontaneous. At the time our observations were made the patient was approximately oriented in all spheres, but displayed a gross defect of recent memory and a long, though somewhat ill-defined retrograde amnesia. The scope of the latter embraced at least ten years of the patient's life. Insight was just beginning to return, but the patient entirely failed to appreciate the full gravity of her memory defect. She showed the usual dullness and lack of spontaneity of the Korsakoff case.

It was unfortunately not possible to obtain a full previous history in this case. It was ascertained, however, that the patient had had an ordinary elementary education and had then taken a full secretarial training course. She had worked for many years as a secretary with various firms, at one time holding a reasonably responsible clerical position in a bank. She had artistic and literary tastes, together with some slight talent for sketching. Intelligence as measured by tests indicated that the basic level was well above average, but there had been considerable deterioration of the standard organic type. By temperament, the patient was reserved and seclusive, but there were no indications of paranoid or obsessional features in the previous personality.

The diagnosis was *alcoholic Korsakoff psychosis*, with polyneuritis.

Analysis of Retrograde Amnesia.—In view of the fragmentary history it was impossible to reconstruct with any confidence the actual course of the patient's life during the years preceding the onset of the psychosis. It appeared plain, however, that remote memory (i.e. for events of childhood and early adult life) was intact,

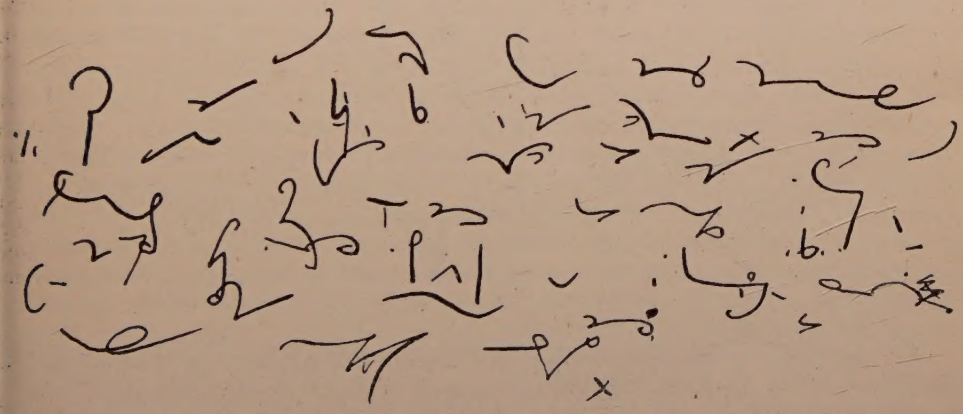
* I wish to express my sincere gratitude to Dr. L. H. Wootton, formerly Medical Superintendent, St. Ebba's Hospital, for his kindness in permitting me to study this patient.

although the chronology appeared somewhat confused. Outstanding events in the patient's personal life up to about the end of the 1914–18 war were accurately recalled, but there was marked uncertainty regarding her movements in the early post-war years. Amnesia appeared to be essentially complete for the fifteen years preceding the onset of the illness, but it was impossible to define its scope with complete conviction.

The prolonged retrograde amnesia, and the lack of full insight into it, gave rise to characteristic distortions of judgment. Thus the patient consistently under-estimated her age by some 10 or 11 years despite the fact that she knew the actual year and could recall her year of birth. She was never fully convinced by arithmetical demonstration of her true age. She was also prone grossly to under-estimate the time that had elapsed since events had taken place which she was still able to recall. Thus she would regularly say that it was only 3 or 4 years since the Armistice. On the other hand, her knowledge of current events was not wholly at fault. Thus she believed that the present King was the "Duke of York" and she had some vague appreciation of the fact that the country was again at war with Germany.

It was of some interest to inquire whether the patient had retained her knowledge of *shorthand* and some dictation tests were accordingly given. One sample of her work is reporduced in Figure 1. The text of this passage reads as follows: "She called upon every women's organization throughout the country to devote itself to war work. Women, she added, were not being fully employed in the war effort. For instance, there were no women on the management committee of the twenty-seven childrens' camps set up under the evacuation scheme, though there are many women with the necessary managerial experience." An experienced stenographer to whom this sample was submitted reported as follows: "I should say that this was the work of a comparative beginner at shorthand, because many of the vowel sounds are put in which may later be left out in practice." Although it is possible that this feature indicates some regression to an earlier level of performance, it must be borne in mind that the patient had not been actively engaged in writing shorthand for some years. In any case, she had learnt shorthand as a girl and could readily remember doing so. The case is not, therefore, parallel to that reported by Stern (1938, p. 252), in which the patient ". . . could write shorthand, which he had learned during the forgotten period, with undiminished skill."

FIGURE 1

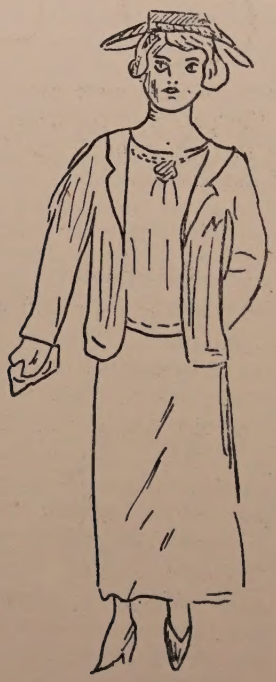


The patient's capacity for *sketching* appeared to have undergone little or no deterioration, but it was not possible to gain access to work which she had executed prior to her admission. The general level of her free-hand drawings was very much above the average.

Observations on Generic Images.—A number of simple tests were given with a view to sampling the patient's generic imagery. Some of the more important of these may be briefly described.

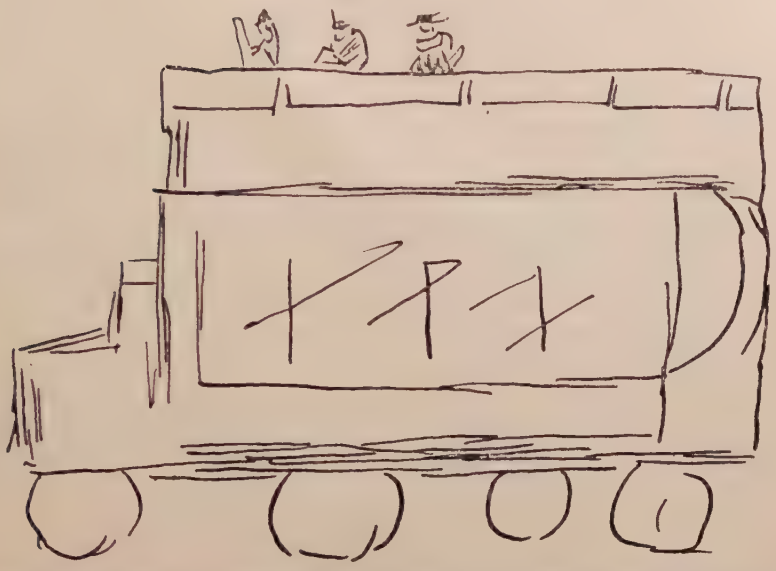
Test 1.—The patient was asked to think of a woman such as she might see any day in the street outside her home. She was then asked to draw a woman such as the one she had just imagined. She was particularly asked to represent the *latest fashion in dress* which she could remember. One of her drawings, which is typical of many obtained under these conditions, is shown in Figure 2. On this occasion, as on all others, the patient was entirely satisfied that *the style of dress represented in the picture corresponded in all essentials with the latest style she could remember and that this represented the "latest fashion," i.e. that of the year of her admission to hospital* (1939). The drawing shown in Figure 2 was submitted to a number of normal subjects of the same sex, and of approximately the same age, as the patient, and the style of dress depicted was on no occasion said to be more recent than 1930. Indeed, most judges dated it somewhere between 1917 and 1924.

FIGURE 2



Test 2.—The patient, who was a Londoner, was asked first to visualize, and then to draw, a bus of the type habitually seen in London during the year preceding her admission to hospital. It was found that she would invariably draw a double-decker bus *with an open top* and alleged that *all double-decker buses presented this appearance.* It will be borne in mind that open-top buses were becoming obsolete in London in the late 1920's and were finally withdrawn from general circulation in the early 1930's. One of the patient's drawings of a bus is shown in Figure 3.

FIGURE 3



Various other tests of the same kind were given with broadly similar results. In general, it appeared that the patient's recollection of objects which had been subject to progressive change in fashion or design had undergone a regression in accordance with the scope of her amnesia.

III

Discussion

It has long been appreciated that the Korsakoff patient regularly interprets his *milieu* in terms of an environment familiar to him at an earlier stage of his life. Although, in the present case, orientation was largely restored by the time the patient came under our observation, there can be little doubt that her outlook remained markedly influenced by the scope of the amnesia. As we have said, it appeared to her to be no more than 3 or 4 years since the end of the Great War and she consistently (and without conscious intention) under-estimated her age by at least 10 years. In much the same way it would appear that her generic images of classes of objects or individuals were, in some cases at least, appropriate to an earlier period of her personal history. She appeared unable to call up generic images built up within the previous 10 years and presumably based on the integration of more recent experiences. It is probable, too, that memory images of individuals displayed the same regressive character as those of more general content. Although no decisive evidence could be obtained in the present case, observations on similar cases certainly point to this conclusion. In short, we may tentatively conclude that *memory images, in their generic or "cumulative" aspects, are subject to the law of regression.*

One may suggest that generic images in the normal individual are constructions based upon a large number of discrete past experiences. If we care, with Bartlett (1932), to employ the term *schema* to denote an active organization of related past experiences, we may say that the generic image is a function of the schema operating *en masse*. But it is plain that more recent experiences contribute predominantly to the content of the image. In the normal individual, indeed, recall of an earlier

and now out-dated fashion in dress or design is notoriously exacting. It is likely, therefore, that our patient, prior to the onset of her illness, would have found it extremely difficult to represent a woman in the fashions of 15 years before. One may suggest that, as an indirect consequence of the amnesia, there is a *release* of memory dispositions (or schemata) built up at an earlier period of life. Indeed, it has long been suggested that past experiences at the threshold of an amnesic gap in the patient's personal memory exercise a prepotent influence in determining the content of disorientation. There is a type of fixation upon an earlier period in the patient's life (Meggendorfer, 1928). Regression in the content of generic images has, therefore, both a negative and a positive aspect. The negative aspect is represented in the failure of recent memory; the positive aspect in the release of schemata normally superseded and held in check by the impact of current events. In consequence, the generic images evoked tend to be in keeping with an earlier era in the psychological history of the patient.*

* The finding that generic images are subject to retrograde amnesia may account for part, at least, of the bewilderment which Korsakoff cases typically display in their dealings with the environment. The confusion attendant upon a lack of correspondence between individual objects and the appropriate generic image may readily be imagined.

IV

References

1. Bartlett, F. C. (1932). *Remembering*. Cambridge.
2. Galton, Sir Francis (1907). *Inquiries into Human Faculty*. Everyman's Edition. London.
3. James, W. (1890). *Principles of Psychology*. London.
4. Katz, D. (1948). Paper read to 12th International Congress of Psychology, Edinburgh. (Unpublished.)
5. Meggendorfer, F. (1928). "Intoxikationspsychosen" in Bumke, O. *Hdbh. d. Geisteskr.*, **7**, 272–85.
6. Ribot, Th. (1885). *Diseases of Memory*. London.
7. Stern, W. (1938). *General Psychology*. New York.
8. Stout, G. F. (1896). *Analytic Psychology*. London.
9. Ward, J. (1918). *Psychological Principles*. Cambridge.
10. Zangwill, O. L. (1941). On a peculiarity of recognition in three cases of Korsakow's psychosis. *Brit. J. Psychol.*, **31**, 230–48.

(*Manuscript received 10th June*, 1949.)

STUDIES IN SPACED AND MASSED LEARNING: III. PAIRED-ASSOCIATE AND SERIAL LEARNING

BY

J. C. TSAO

(*From the University of Hong Kong*)

An experiment on paired-associate learning was carried out to test Hovland's prediction hat spaced practice should be more advantageous than massed if the order of presenting he pairs of nonsense syllables were fixed.

In the results, spaced and massed practice showed little difference in learning efficiency. Iovland's prediction was not confirmed.

In a previous experiment of this series of studies it had been found that spaced practice vas more efficient than massed in serial learning when syllables of "low-meaningfulness" alue were used; but that there was little difference in efficiency between the two forms f distribution when "high-meaningfulness" syllables were used. The results of the two xperiments naturally suggested that, as regards the relative efficiency of spaced and nassed practice, paired-associate learning of low-meaningfulness syllables differed from erial learning of similar syllables, but matched serial learning of high-meaningfulness yllables. It seems justifiable to draw the conclusion that the relative efficiency of spaced nd massed practice is different in paired-associate and serial learning, not because remote ssociations can be formed in the one case and not in the other, as Hovland (1939*a*, 1939*b*) ssumed; but because the organization of the material and the learning processes are ifferent in the two forms of learning.

I

Introduction

N two experiments, Hovland (1939*a*, 1939*b*) found that reminiscence was very narked, and distributed practice was superior to massed practice in serial learning; ut that this was not so in learning paired associates which were presented in a onstantly changing order. Hovland explained the results by assuming that fewer emote associations could be formed with paired associates and that consequently here was less inhibition, and thus reminiscence effects and the advantage of spacing, oth of which depended on the dissipation of inhibition during interpolated intervals, ere naturally less conspicuous in paired-associate than in serial learning. He redicted that, if the order of presentation of the successive pairs was not varied, eminiscence effects and, by implication, the advantage of spaced practice, would be ntermediate between his results and those in serial learning.

McClelland (1942), and even Hovland (1939*a*) himself, suspected that the absence f reminiscence and the small advantage of spacing in Hovland's results might be ue to the exposure time which was 2 seconds for each member of a pair. Neither Iovland nor McClelland found reminiscence at a 4-second rate of presentation. Iovland, however, thought that this result was unlikely, and McClelland did not ubject his hypothesis to test.

Hovland's prediction concerned the order of presentation, fixed or varied; and IcClelland suspected the effects of 4- or 2-second exposure time. The "prediction" nd the "suspicion," put together, lead to an interesting problem: if paired associates ere learned with a 2-second exposure time for each pair and a fixed order of presenting he successive pairs, would spaced practice be much more efficient than massed, ust as in serial learning? In order to answer this question the present experiment as carried out.

II

Method

Material.—Two lists of eleven pairs of nonsense syllables each were used. The syllables were selected from Hull's (1933) list; all of them were of "meaningfulness value" from 0 to 5 on Hull's scale. To each list an additional pair was attached, which served as an example. The two lists are as follows:

LIF KUJ	LED HIJ
KUJ	LED
WIB FOV	BIW ZUD
HAJ YUT	VEF ZOJ
JID VAK	PIJ TEV
WOJ FEP	MUJ VIP
ZIH VOF	WOF ZIK
VAW KOY	KEF HUJ
ZUX MIV	YOJ VUK
FUP BEJ	FAP YUF
VEP BUW	POB MAF
ZOY TIV	TUJ VAB
KEJ ZIB	KIH MEV
FOV	BIW
HAJ	VEF
JID	PIJ
FEP	VIP
VOF	WOF
VAW	HUJ
MIV	VUK
FUP	FAP
BUW	MAF
TIV	TUJ
KEJ	MEV

Subjects.—The subjects were twenty-four men, twenty of whom were R.A.F. personnel and four R.N. staff. None of them had any previous experience of learning nonsense syllables or knew the purpose of the study before or during the experiment.

Procedure.—The general procedure was similar to that described in the first article of this series (Tsao, 1948*a*). The same memory apparatus was used. The "example" pair of the first list was first exposed and explained to the subject; the eleven pairs followed the example. Each pair was exposed for 2 seconds with an interval of $\frac{1}{2}$ second between each successive pair. The single syllables followed the eleventh pair, and the subject was asked, on seeing a single syllable, to pronounce the other one which had been paired with it. Ten trials were given of each list. The order of pairs and single syllables was the same throughout the experiment as shown in the lists given above. In spaced practice, a 1-minute interval was interpolated between successive trials, during which the experimenter talked with the subject on current events or other general topics. In massed learning the readings were continuous. In order to balance practice effects and the influence of shifting from spaced to massed practice or vice versa, the learning of the two lists was arranged in such a way that each was learned, by different subjects, for an equal number of times in every possible situation in relation to each other and to the distribution of practice.

III

Results

TABLE I

Scores (Mean Number of Correct Responses) in Paired-Associate Learning by Spaced and Massed Practic

Practice	Trials									
	1	2	3	4	5	6	7	8	9	10
Spaced ..	0·46	0·75	0·96	1·46	1·46	1·67	2·21	2·17	2·54	2·92
Massed ..	0·42	0·83	1·00	1·21	1·25	1·38	1·75	2·17	2·25	2·71

The learning score of each subject in each trial is the number of correct responses. Table I contains the average scores of each trial in spaced and massed practice. Since there were twenty-four subjects and each of them learned once by spaced practice and once by massed practice, each average score is based on twenty-four individual scores. The results are also presented in a graphic form in Figure 1.

FIGURE 1

Learning Curves of Spaced and Massed Practice in Paired-Associate Learning

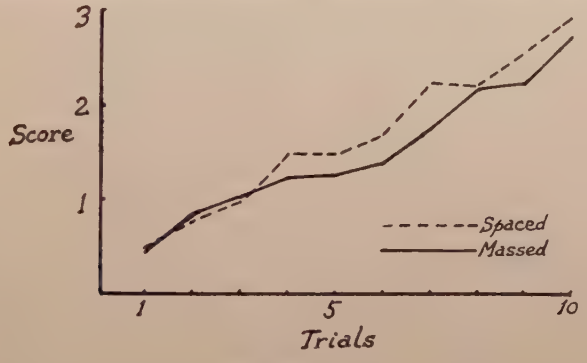


TABLE II

Scores of, and Differences Between, Spaced and Massed Practice in Paired-Associate Learning

	Spaced		Massed
10th Trial:			
M.	2·92		2·71
Dif.		0·21	
C.R.		0·44	
9–10 Trials:			
M.	5·46		4·96
Dif.		0·50	
C.R.		0·55	
1–10 Trials:			
M.	16·58		14·96
Dif.		1·62	
C.R.		0·53	

From the data in Table I and the curves in Figure 1, it can be seen that the efficiency of massed practice is generally on a lower level than that of the spaced learning, but the discrepancy between these two forms of practice is small. The differences, and the corresponding C.R.'s, between the spaced and the massed practice in three measures, viz. mean scores of the 10th, 9–10th and all trials, are given in Table II. None of the differences is statistically significant. The results show, therefore, that in paired-associate learning spaced and massed practice made very little or no difference in learning efficiency, or more concretely, that 1-minute intervals interpolated between successive trials did not greatly affect learning efficiency in paired-associate learning.

IV

Discussion

The results of the present experiment are quite different from those of serial learning of low-meaningfulness syllables in another experiment of this series (Tsao, 1948*b*), which showed that spaced practice was much more efficient than massed. This may be an indication that there is some difference between paired-associate and serial learning, which caused the difference in efficiency of distributions of practice in the two types of learning.

The technique used in the present experiment was different from Hovland's (1939*a*) in the following aspects: (1) In the present experiment the order of pairs in presentation was fixed while Hovland constantly changed it. (2) The exposure time for each pair was 2 seconds in this experiment while in Hovland's it was 4 seconds. (3) The two members of each pair were exposed simultaneously inside the same window in the present experiment, while in Hovland's the two members were shown successively and alternatively in two windows on his apparatus. Hovland predicted that a constant order of presentation of pairs would induce remote associations and render spaced practice advantageous because of the dissipation of inhibition during interpolated intervals. Both Hovland and McClelland suspected the possibility that, with a 2-second exposure time for each pair, reminiscence effects and, by implication, the advantage of spaced practice, would be found. Hovland (1939*a*) thought that, in his experiment, "the continuous rotation of the order of the pairs and the alternative exposure of the syllables in different windows of the apparatus is less monotonous to the subject, so that decremental effects are less in paired-associate lists than in serial learning where the syllables are presented in the same order, in the same window and at the same rate." In fact, the results of the present study do not confirm any of Hovland's or McClelland's suppositions.

Hovland assumed that the difference between paired-associate and serial learning lay in the formation of remote associations. Remote associations, according to Hovland, were usually formed abundantly in serial learning, but not in paired-associate learning in which the order of presentation of the pairs varied constantly. Time intervals, during which inhibitions due to remote associations dissipated, were, therefore, more beneficial in serial than in paired-associate learning. In paired-associate learning with a fixed order of presentation of the pairs, Hovland predicted remote associations would be formed. This part of Hovland's prediction was right; it was noticed in the present experiment that the subject responded very frequently to a single syllable with a syllable other than its partner in the list, sometimes even with a syllable in the other list. Hovland's prediction broke down, however, in the

other part; spaced practice was not much superior to massed in paired-associate learning with a fixed order of presentation of the pairs, although plenty of remote associations had been formed, as is shown by the results of the present experiment.

With the failure to confirm Hovland's prediction, his hypothesis of inhibition apparently cannot be used to explain the fact that spaced practice was much more efficient than massed in serial learning of low-meaningfulness syllables, but was not so in learning paired associates of similar material as shown in the present and the other experiment (Tsao, 1948*b*). The differential effects of distribution of practice may, instead, be caused by some differences both in the organization of material and in the learning processes of the two forms of rote learning.

In serial learning a list is just a chain of nonsense syllables. Its items, at the beginning, are homogeneous to the subject. This, as noticed by Restorff (1933), makes serial learning a difficult task. Only as the learning proceeds, grouping and other devices, as observed by Müller and Schumann (1894) and later investigators, are developed by the subject, which make the syllables more accessible and, furthermore, facilitate the learning. Paired associates, on the other hand, provide some organization in the material. From the very beginning the syllables are given in pairs which may be taken as some sort of "sub-wholes." Since each syllable in paired associates is "matched" by another one and associated with it, it is thus given an "artificial meaning." So the material in paired-associate learning may be regarded as better organized and even a little more "meaningful" than the material in serial learning. Meaningful materials are usually learned about equally well with spaced and massed practice, as has been shown in a former experiment of this series (Tsao, 1948*b*).

The above refers to the organization of the material; there are also some differences in the learning processes in paired-associate and serial learning. Both in paired-associate and serial learning a syllable has to be memorized as a single syllable. Also, in serial learning a syllable has to be memorized in its serial position, that is, before and after certain syllables, especially after a certain syllable. In paired-associate learning, a syllable has to be remembered with a certain partner which forms a pair with it and serves as the cue for its recall. In serial learning, with the anticipation method, the subject must go back to the preceding syllable in order to give a correct response. He has to remember the preceding one as well as the very syllable concerned, before he can anticipate it correctly. Moreover, a syllable is remembered not only as a syllable itself, but also as a cue for recalling the following one which is still to come. The syllables are exposed individually, but the subject sees only one syllable at a time, and he must relate it to both the preceding one which is no longer present and the succeeding one which is not yet in view. Serial learning is, therefore, a rather complex process. Paired-associate learning is different. A pair of syllables is shown at a time. In testing the memory, one of them is given; what the subject has to do is only to recall the other one. Since in learning both syllables are exposed simultaneously, the subject does not have to associate one of them with something not in view. In comparison with those of serial learning, the processes in paired-associate learning are much simpler. A more complicated learning process is, of course, a more difficult one, and thus serial learning and paired-associate learning differ in difficulty. That the advantage of distribution of practice is more prominent in difficult tasks than in easy tasks has been proved by Cook (1928), Borovski and Kobozeva (1935) and many others.

This study was carried out in the Psychological Laboratory, Cambridge. I wish to express my gratitude to Mr. G. C. Grindley for help and advice.

V

References

1. Borovski, V. M., and Kobozeva, P. S. (1935). Relation between distribution of practice and difficulty of the problem. *Instinkt. Naviki.*, **1**, 68–76.
2. Cook, S. A. (1928). The effect of various temporal arrangement of practice on the mastery of an animal maze of moderate difficulty. *Arch. Psychol.*, **15**, *No.* 98.
3. Hovland, C. I. (1939*a*). Experimental studies in rote learning theory: IV. Comparison of reminiscence in serial and paired-associate learning. *J. Exp. Psychol.*, **24**, 466–484.
4. Hovland, C. I. (1939*b*). Experimental studies in rote learning theory: V. Comparison of distribution of practice in serial and paired-associate learning. *J. Exp. Psychol.*, **25**, 622–633.
5. Hull, C. L. (1933). The meaningfulness of 320 selected nonsense syllables. *Amer. J. Psychol.*, **45**, 730–734.
6. McClelland, D. C. (1942). Studies in serial verbal discrimination learning: I. Reminiscence with two speeds of pair presentation. *J. Exp. Psychol.*, **31**, 44–56.
7. Müller, G. E., and Schumann, F. (1894). Experimentelle Beiträge zur Untersuchung des Gedaschtnisses. *Zeit. Psychol. Ergbd.*, **1**.
8. Restorff, H. von (1933). Uber die Wirkung von Bereichsbildung im Spurenfeld. *Psychol. Forsch.*, **18**, 299–342.
9. Tsao, J. C. (1948*a*). Studies in spaced and massed learning: I. Time period and amount of practice. *Quart. J. Exp. Psychol.*, **1**, 29–36.
10. Tsao, J. C. (1948*b*). Studies in spaced and massed learning: II. Meaningfulness of material and distribution of practice. *Quart. J. Exp. Psychol.*, **1**, 79–84.

(*Manuscript received 20th August*, 1948.)

THE EFFECTS OF REPEATEDLY RETESTING THE SAME GROUP ON THE SAME INTELLIGENCE TEST: II. HIGH GRADE MENTAL DEFECTIVES

BY

A. W. HEIM and J. G. WALLACE

(*From the Medical Research Council Applied Psychology Unit, Cambridge*)

Twelve mentally-defective schoolboys between 14 and 16 years of age took an ntelligence test (AH 4) once a week, ten times. They took also Progressive Matrices nce, after they had finished the AH 4 testing. Subjects of low intelligence were chosen n order to throw some light on the findings of a previous experiment on repeated retesting, n which the subjects were intelligent adults. The main aims of the schoolboy enquiry vere to ascertain how far the flattening of the progress curves observed in the earlier nquiry was due to the artificial "ceiling" imposed and how far the part played by variation vithin the individuals varies with the level of intelligence of the group.

The mentally-defective subjects showed a gradual improvement in test performance vith little sign of flattening towards the end; their individual progress curves were far ess smooth than had been those of the more intelligent subjects. Other findings included elatively high proficiency on the diagrammatic part of the test, as opposed to the verbal nd numerical, and some transfer of training from AH 4 to Progressive Matrices.

I

Introduction

N a previous report (Heim and Wallace, 1949) the results of repeatedly retesting small group of intelligent adults once a week, for some weeks, were described. t was found that the typical individual progress curve showed a steep upward rise or the first four or five testings, followed by a levelling out towards the end. The rder of the subjects remained strikingly constant, but the correlations between he two parts of the test tended to fall with successive testings, suggesting that the art played by "bias" increases with repeated testings. There was some evidence hat Part I (verbal and numerical) presented less difficulty to the group than did art II (diagrammatic).

These findings raised questions of theoretical and practical interest. Does ntelligence test performance, like other skills, always improve with practice? Does ate of improvement always become negligible after the first few testings? Or are hese findings dependent upon a particular relation between group and test used?—.e. a pre-selected group whose intelligence level is higher than that of the population or whom the test was originally designed. Does the "learning" extend to other omewhat dissimilar tests? Can subjects within the group always be relied on to naintain their relative position or is this, again, dependent on the relationship etween group and test used?

The solution of these and allied problems requires a series of experiments with ifferent groups and different tests. Mental defectives were chosen in this instance, ince it was felt that they would have the greatest scope for improvement: there vould be little danger of their reaching the end of the test and gaining nearly full narks, even after many testings.

II

Subjects and Tests

The subjects were twelve high-grade mentally-defective boys, ranging in age from 14·6 to 15·10, and drawn from Littleton House School, Cambridge. Only boys who were competent readers were accepted as subjects, in view of the largely verbal bias of the first part of Test AH 4. It was clear that any Part I scores gained by boys who have difficulty in reading simple words would be worthless.

The Binet I.Q.'s of the boys, established before they were admitted to the school, ranged from 55 to 76·5. The period between the Binet testing and the AH 4 testing varied from one to five years.

AH 4 testing was carried out weekly, for eleven consecutive weeks, at the end of which period the school term came to an end. In this way eight of the boys were tested eleven times and all twelve of them were tested ten times. Thus in four cases only there was on some one occasion a gap of two weeks between testings. The results discussed in this report are concerned, unless otherwise stated, with the ten tests taken by the whole group of twelve.

Every care was taken to ensure that the subjects understood what was required of them: the usual simple examples were given of the type of question met with in both parts of the test and the boys worked through these in their own time, each week, before embarking on the test proper, each part of which lasts 10 minutes.

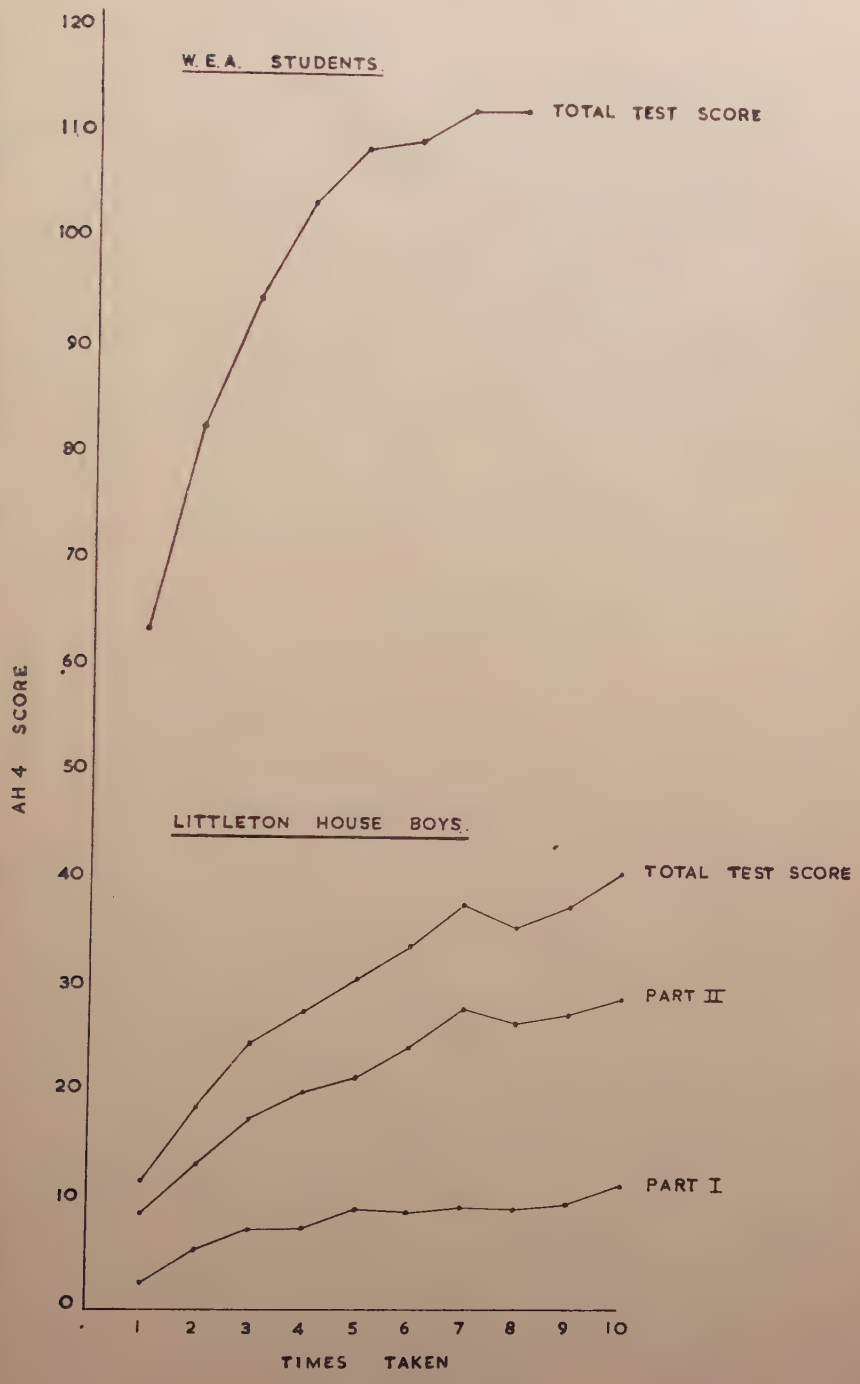
The concentration of the mental defectives during testing was, of course, markedly inferior to that of the W.E.A. students who had taken part in the first enquiry. However, the boys appeared less subject to boredom in the course of testings—perhaps because they, unlike their predecessors, were always meeting new questions as they got further in the test throughout the ten weeks, perhaps because boredom with familiar things is a characteristic of high rather than low intelligence. The boys were interested and forthcoming; they were, for instance, very eager to see the stop-watch at the end of each testing and, if possible, to hold it. They were quite keen to hear whether they were improving as a group, and when they were told that they had done the test for the last time, they looked downcast and asked the tester when she would be coming back. The eleventh testing took place on 1st April, 1947. On that occasion several of the boys gleefully made mistakes over the Examples, in order to have the pleasure of making an "April Fool" of the tester.

At the end of April, after their school holiday, the twelve boys were given the Progressive Matrices Test. Two unusual points of procedure should be mentioned in connection with the latter testing. (*a*) Six examples of "Matrix" type devised by the tester were provided to be worked through, before the test proper was begun. The first and fifth of these had already been completed, but the boys were urged to study them; the remaining four examples had to be solved by the boys themselves. (*b*) A time limit of 30 minutes was imposed for the test proper. Five of the boys finished the test within this time and all but one of the others had reached set D (the fourth of the five sets of twelve questions).

In addition to the Binet I.Q.'s and the Progressive Matrices scores, the writers had access to the boys' school reports, which include remarks on performance in school subjects (including manual ones), speech, physique, behaviour and family history. The Headmaster kindly added his comments on the appropriateness, in his view, of the Binet I.Q. with which the boy had entered the school, and he also arranged the subjects in rank order according to his conception of their "innate all-round ability." The comparison of these criteria one with another yielded little of direct relevance to this particular enquiry. In order to save space, therefore, discussion of these results has been omitted from this paper.

FIGURE 1

PROGRESS CURVES OF MEAN SCORES



III

Results

A. The General Trend of Scores.

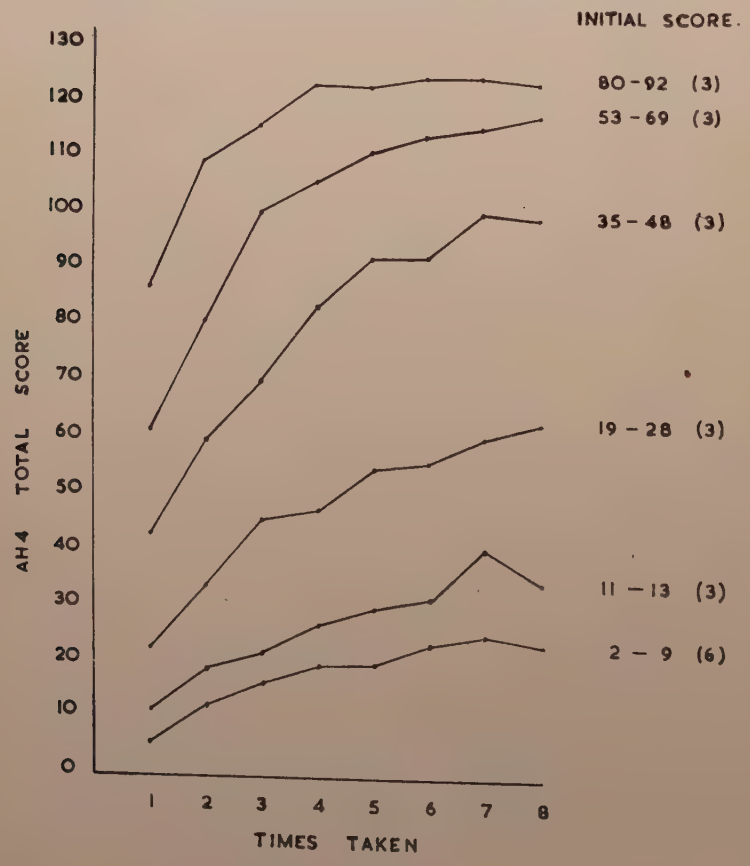
Figure 1 shows the curves for the mean scores of the group of twelve subjects, for the ten testings. Separate curves are shown for Part I, Part II and total test scores. The total test score curve for the W.E.A. group is shown above.

It may be seen that the total score curves of the Littleton House boys and of the W.E.A. students differ considerably from each other. The W.E.A. curve rises steeply at first and begins levelling out somewhere around the fifth testing. The mentally-defective group, however, shows a fairly steady rise—except for trial 7 which yields a small peak.

This would suggest that the flattening out of the W.E.A. curve is a property of the relation between the test (intended for a cross-section of the population) and the group (a selected, intelligent group), rather than of the learning process as such. Given this relatively easy task, the W.E.A. group virtually finished all its "learning" after some five trials. The schoolboys' curve suggests that the duller subjects learn not only more slowly, but also over a considerably longer period, given the opportunity.

FIGURE 2

Progress Curves Over 8 Trials at Different Levels



That the difference in the trends is a difference in degree rather than in kind is suggested by Figure 2 (for which idea we are indebted to Mrs. N. Harris). In this figure the results of the subjects from both groups over eight trials are included: each progress curve shows the mean for three subjects—adjacent on their first score—except for the bottom progress curve which is the mean of the *six* lowest-scoring subjects.

This figure illustrates clearly the change in gradient of the progress curve as the intelligence of the subjects decreases. It is observable that the "steps" taken by the brighter groups grow progressively smaller with each testing, whereas this is not true of the duller groups.

Inspection of the schoolboys' progress curves for the two separate parts of the test (Figure 1) throws some light on what is happening. The Part II scores are initially, and throughout subsequent testings, much higher than those for Part I; moreover, the improvement on Part II is far more marked than that on Part I. Linear regressions were fitted to the curves of Part I, Part II, and also of Part II minus Part I, in order to ascertain the improvement on Part II relative to Part I: these showed that the improvement is significantly greater for the second Part.

Thus it is clear that Part I contributes little to the improvement shown throughout the Total score curve: the shape of the Total score curve is very largely determined by that of the Part II curve. Evidently Part I is almost outside the intellectual range of the mental defectives. It appears unlikely that they would ever make a fair showing on Part I, even if they continued to take it for months, or if they tried taking it with unlimited time. On the other hand, their Part II scores suggest that they might well continue to improve if they went on taking Part II.

The initially higher scores and the greater progress made in Part II lends itself to various interpretations. It may be that unintelligent subjects tend to find diagrams intrinsically easier to deal with than words or figures. A more direct approach is possible with the diagrams: they represent nothing beyond themselves. But words and figures are one degree removed from what they represent; they are symbols to be read and translated before they can be used. It may be that at low levels of intelligence these processes present a real difficulty. Beyond a certain intelligence level this distinction evidently plays a far smaller part: intelligent subjects vary as to which "bias" they prefer.

A possible explanation of the difference in *slope* between the progress curves of the two Parts may lie in the fact that the subjects have had practice in school for some years, in dealing with words and figures, whilst they have had little or none with diagrams of the kind used in the test. They may therefore have made most of the progress of which they are capable in the first field, whereas they have considerably more scope—and, perhaps, interest—in the second.

That the difference between the Part I and Part II progress curve is not a simple matter, is evident from a comparison of the number of questions attempted with the ratio of number correct to number attempted, on the two Parts.

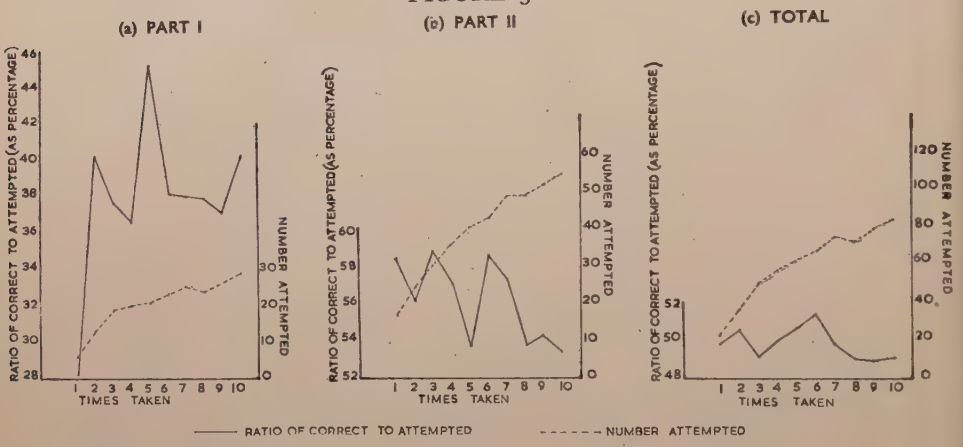
Figures 3 (*a*), (*b*) and (*c*) show these two curves for the two separate Parts of the test and also for the Total test score. It may be seen in Figures 3 (*a*) and (*b*) that, whilst the number of questions attempted increases to a greater extent and more rapidly for Part II than for Part I, the ratio of correct answers to number of questions attempted tends to decrease in Part II and, if anything, to increase in Part I—in so far as such serrated curves may be said to show a tendency.

This may be due to the fact that the questions in the test become progressively harder: in Part II the quicker boys were reaching quite difficult questions during

the last few testings, whereas none of them got very far in Part I. Furthermore, the Part II ratio was startlingly higher to begin with than the Part I ratio. The difference in trend between the two ratios might, of course, be due, to some extent, to the greater practice which the boys got on Part I, owing to their always attempting the *same* questions week after week.

On the other hand, for the W.E.A. subjects, the curve showing the number of questions attempted was similar in shape to their test score progress curve. But the ratio curves for the two Parts were diametrically opposed; in Part I the ratio sank steadily, and in Part II the ratio (*lower*, initially, than that for Part I) rose steadily—though it was never as high as the ratio for Part I. Thus the brighter subjects (with their somewhat verbally biassed occupations) evidently found Part I absolutely, as well as relatively, easier than the diagrammatic second Part.

FIGURE 3



B. Individual Variation.

The scores on each testing were compared with the scores on every other testing and also with the "pool" of the scores. In correlating each set of scores with the pool, the set in question was included in the pool. The product moment method of correlation was used in this investigation as in the earlier one. It was chosen in preference to the method of rank correlation in order that occasional striking differences between particular individuals should not be masked and that a measure of the significance of differences would be available.

Using the scores of the whole Littleton House group, the test yielded very high consistency. The correlations between the separate testings and the "pool" ranged from 0·879 (second with pool) to 0·989 (fifth with pool). The correlations between scores on the different testings with one another were scarcely less close: with one exception (the second with the seventh, 0·709) they were all over 0·80, the majority exceeding 0·90.

In addition to the above comparisons between Total test scores, scores on the separate Parts of the test for each testing were correlated with their pool. These correlations were found to be very similar to those obtained with Total test score: they were all very high (over 0·785). No significant difference was found between the two Parts, and no trend was observable with repeated testings. Thus, all the correlations *between the testings* are highly significant, none of the differences *between the correlation coefficients* being significant.

Figure 4 (*a*) shows the individual progress curves of the twelve boys, numbered for purposes of reference. Inspection of these curves reveals that they fall fairly clearly into two separate groups, Nos. 1–4 in the first group and Nos. 5–12 in the second; and that the progress curves of the second group in particular cross and re-cross with great frequency. It was thought interesting, therefore, to examine the consistency of the mental defectives within these two groups. The correlation coefficients showed a very changed pattern; the range for the total group of twelve was $r = 0{\cdot}709$ to $0{\cdot}989$ (a difference of $0{\cdot}28$), whereas that for the lowest eight was $r = -0{\cdot}008$ to $0{\cdot}917$ (a difference of $0{\cdot}925$); only twenty out of the fifty-five correlation coefficients on the smaller group were significant; and a strong trend towards greater consistency was shown *with later testings.*

FIGURE 4 (*a*)
INDIVIDUAL PROGRESS CURVES
LITTLETON HOUSE GROUP

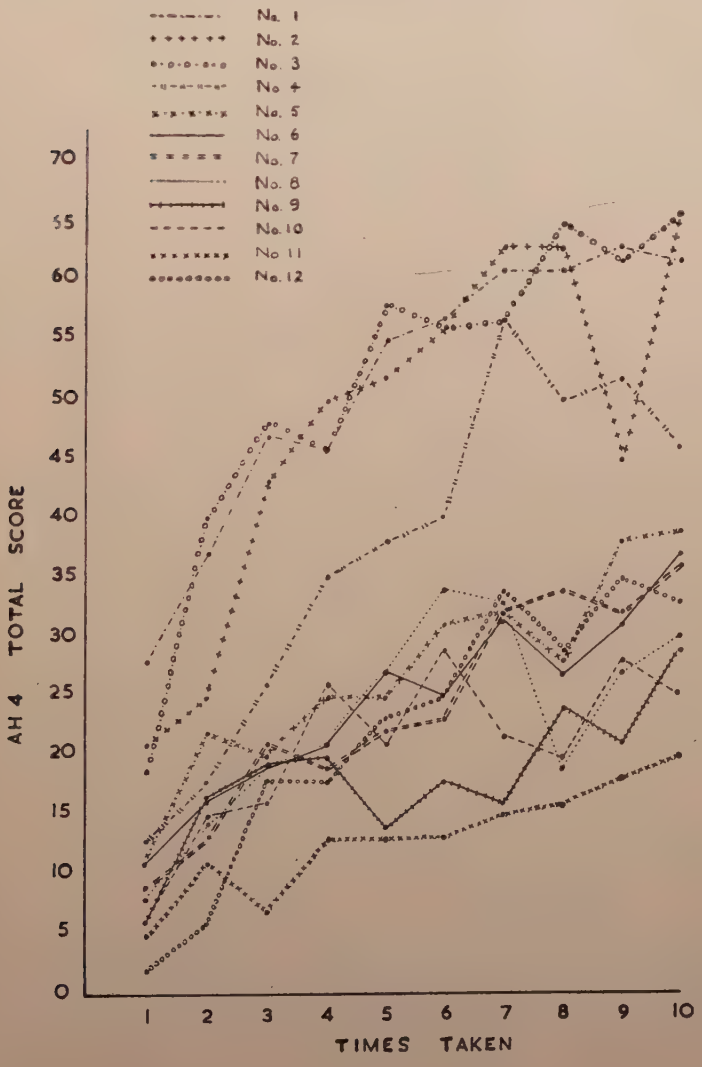
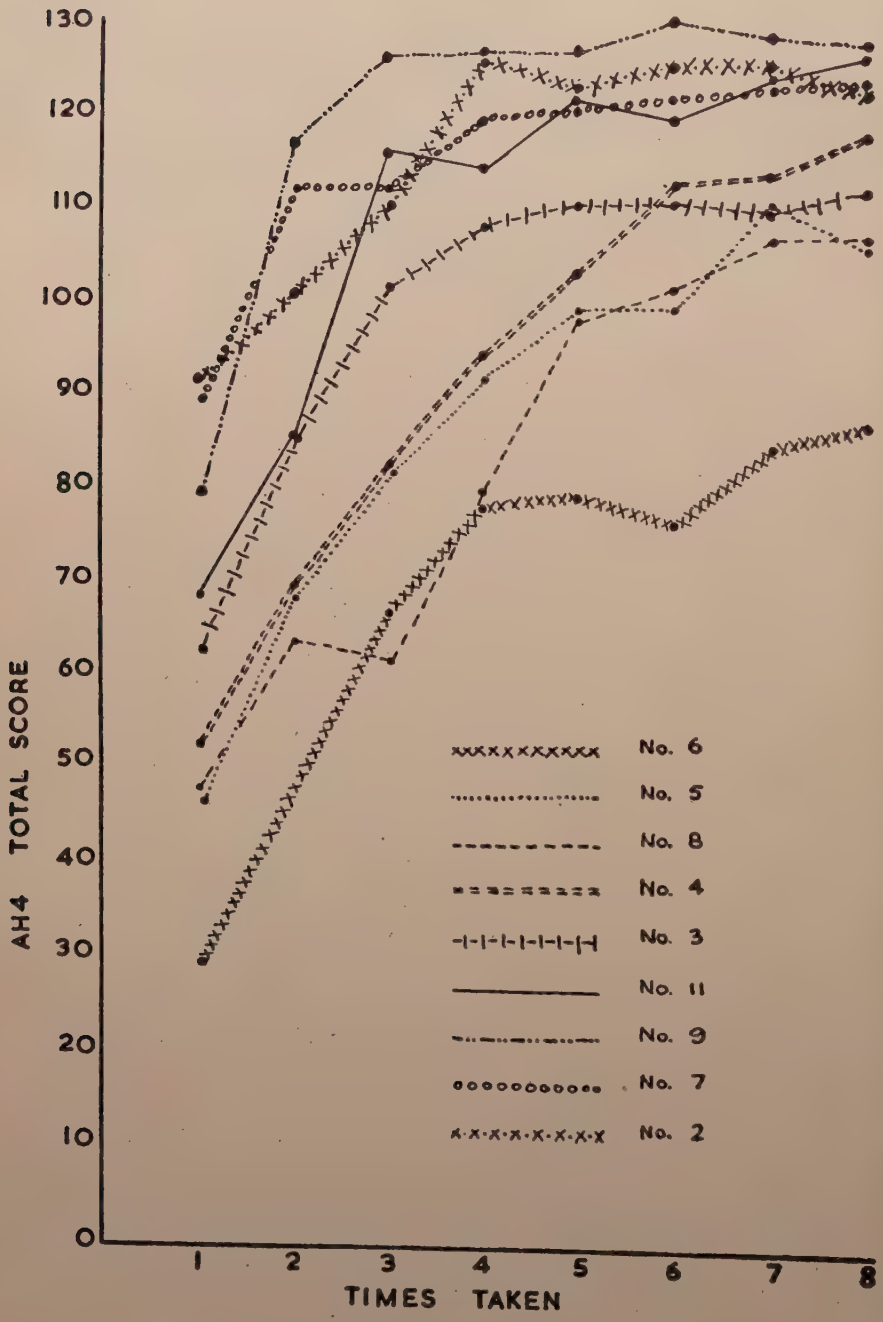


FIGURE 4 (*b*)

INDIVIDUAL PROGRESS CURVES

W.E.A. GROUP



An Analysis of Variance showed that significant differences in ability obtained between the subjects on both Parts of the test. It is clear, however, that the consistency of the total group is due largely to the fact that the first four maintained their position relative to the other eight throughout the experiment.

The relative inconsistency of the dullest eight boys as compared with the total group, and as compared with the W.E.A. group, is to be expected on statistical as well as psychological grounds. Their range of score at each testing is extremely small; differences of one or two marks, such as they exhibited on initial testing, have little or no meaning; and chance played some part in determining each individual's score (see Section III (D)). Mental defectives are notoriously prone to instability and lack of concentration and many of the Littleton House boys were pre-selected for behavioural as well as intellectual difficulties. It is possible that a sufficiently low-grade test would reverse the finding—bunching the brighter subjects so that their inter-test consistency falls and discriminating between the duller ones more adequately and, hence, more consistently.

C. "Learning" in Intelligence Test Taking.

It is clear from Figure 1 that a good deal of "learning" takes place, as a result of repeatedly taking the same intelligence test, without any knowledge of results on the part of the subjects. The speed and extent of the learning varies with the intelligence level of the subjects and the bias of the test, but it seems to be always present in some degree.

The writers thought that it would be interesting to see whether this learning carried over to other intelligence tests. For this reason they gave the Progressive Matrices Test to the Littleton House group, when the AH 4 testing was finished. Progressive Matrices was chosen because it has been widely used in the past with groups of varied intelligence and because it was clear from the AH 4 findings that a test with a visual bias would yield more fruitful results than one with a verbal or numerical bias.

As noted above, examples were provided for the boys to work through at leisure before embarking on the test proper. The examples were easy, but they did illustrate the kind of principle that the subjects would meet in the test and the form of answer required.

The time limit of 30 minutes was imposed for purely practical reasons: the examples took some 15 minutes to complete and it was felt that longer than 45 minutes would probably overtax the boys' attention. In fact, the results showed that this timing was legitimate: the majority of the subjects reached the last section but one of the test and five of these finished the whole test in the 30 minutes. In view of the high proportion of errors made in the last two sections it was felt that the scores would have been substantially the same if unlimited time had been allowed.

None of the subjects had ever seen this test before.

The twelve Matrix scores were as follows: 41, 37, 36, 34, 30, 30, 25, 24, 21, 17, 16, 8. The following table shows two sets of Matrix norms used by the Army in the 1939–45 war. They are based on a sample of 5,000 representative of the war-time Army in terms of age, educational standard and non-verbal intelligence (Col. B. Ungerson, personal communication). The sample is not strictly comparable with the civilian population through the absence of those in Reserved Occupations and those rejected as mentally or physically unfit. The frequencies of the Littleton House scores for each grade are given in columns 4 and 6.

%	Grade	20-minute norms		45-minute norms	
		Matrix marks	*Number of Littleton House subjects*	*Matrix marks*	*Number of Littleton House subjects*
10	A	46+	0	54+	0
20	B	40−45	1	48−53	0
20	C+	35−39	2	44−47	0
20	C−	29−34	3	39−43	1
20	D	20−28	3	29−38	5
10	E	−19	3	−28	6

Thus the general standard is extraordinarily high for a group of subjects whose Binet I.Q.'s are all under 80. It may be seen that only half of the boys come within the bottom 10 per cent. of the Army population on the 45-minute norms, and only one quarter on the 20-minute norms.

In view of the difference in form between Progressive Matrices and even the visual-biassed Part II of AH 4, this suggests some transfer of training. The corresponding grades on AH 4, on the first testing, were D, 2, and E, 10. On a bigger, more representative group, correlations of 0·66, 0·68 and 0·69 (Parts I, II and Total respectively) have been found between AH 4 and Progressive Matrices. It looked, therefore, as though the comparatively high Matrix scores were due largely to the "training" effect produced by repeatedly taking a different test, without knowledge of results.

There was a possibility, however, that these "high" scores were due to the presentation of Matrix-type examples to be correctly completed by the boys before they embarked on the Matrix Test proper. (See Wallace, 1949, Section II, 4, for influence of Examples on test scores.) A control group of Littleton House boys, comparable with respect to age and reading competence, were therefore tested the succeeding term on Progressive Matrices under exactly the same conditions as were the first group—save that the second group had never seen AH 4.

The scores of this group were as follows 40, 32, 27, 27, 23, 14, 14, 8, 7, 6, 6. The mean scores for the two groups were 26·6 for the test group and 18·5 for the control group. This is clearly a substantial difference, being a drop of over 25 per cent. It is not, however, statistically significant, $t = 1{\cdot}8$—instead of the requisite 2·08. These results are not, therefore, wholly conclusive: they suggest that the high Matrix scores of the first group of boys were due partly to the practice effect of repeated AH 4 testings and partly to the inclusion of the examples with the Matrix Test.

The effect of providing examples for the subjects to work through, and of test sophistication, can scarcely be overstressed in view of the increasing employment of psychological tests for practical purposes, and the practice of many psychologists of treating I.Q. as, within narrow limits, a reliable, immutable measure.

D. Differences between Groups of High and of Low Intelligence.

The most striking difference between the W.E.A. students and the mentally-defective schoolboys was the difference in consistency of individuals. Figures 4 (*a*) and (*b*) show the individual progress curves for the two groups.

It may be seen that the high-grade subjects, with very few exceptions, yielded smooth curves which rarely crossed one another. The progress curves of the duller

subjects, however, oscillate wildly and repeatedly cross and re-cross one another. The erratic course of the individual boys is completely obscured in the progress curve of their means (see Figure 1). Most of the individual curves of the W.E.A. students strongly resemble their mean curve, whereas only one of the schoolboys (No. 1) produced a curve at all similar to the Littleton House mean curve. Discussions with the Headmaster, and with the boys after testing, threw no light on the violent changes in score from one week to another. No explanation was found, for instance, for 2's sudden drop at his ninth testing or for 4's impressive rise between sixth and seventh testing.

Inspection of the curves showing the ratio between correct answers and questions attempted showed the same difference between groups. The W.E.A. curves were smooth, they tended to fall for Part I and to rise for Part II. The Littleton House curves are so jagged that little can be deduced from them, but in so far as they exhibit any trends, these go in the opposite direction from those of the W.E.A. group.

The relationship between Parts I and II of AH 4 pursues the same oscillating course for the mental defectives—and the same relatively smooth one for the intelligent subjects, see Figures 5 (*a*) and (*b*). Whilst the smoothness of the W.E.A. curve is broken only twice (at the second and eighth testing), the Littleton House "curve" proceeds in erratic, irresponsible manner, to finish almost exactly where it began.

The general trend, however, is downward, as it is with the W.E.A. group: both the peaks and the dips, looked at separately, become progressively lower. In the W.E.A. report it was suggested that this decrease in relationship between Parts I and II might be due to the narrowing range of scores of the group, the standard deviation having been found to fall from 19·54 the first time, to 12·13 the last time. These new data suggest that this greater selectivity may not have been the main cause, since the standard deviation of the mentally-defective subjects tended to *rise* throughout the testings (first time 7·33, tenth time 15·24), no artificial ceiling being imposed in this instance.

It would appear then that the effect of individual differences in bias tends to increase with retesting, in any case when the test used fully extends the subjects. In view of the erratic nature of the curve in Figure 5 (*a*), however, further data on this point should be acquired, preferably with an unselected group of subjects.

The differences between the mean curves of Part I and Part II scores given in Figure 1, shows how important a rôle bias plays, whether the subjects are test-sophisticated or not, when they are of low intelligence.

The Littleton House subjects frequently omitted questions in Part I though, with the exception of No. 11, they never omitted questions in Part II. The W.E.A. group always took the questions in order, on both Parts of the test.

The number and type of mistake made by the mentally-defective boys was compared with that made by the W.E.A. students. Apart from the fact that the boys made far more mistakes than the students, the most striking difference lay in the consistency of the errors. The students had, with rare exceptions, stuck to the right answer once they had achieved it, and they tended to give the *same* wrong answer to questions they failed to understand. The Littleton House boys, however, would often give a wrong solution to a question they had previously answered correctly, and would very often give a series of wrong answers, week after week—changing from one incorrect solution to another. This would suggest that their results were due entirely to chance, but the Analysis of Variance did suggest some measure of genuine improvement.

FIGURE 5 (*a*)

CORRELATION BETWEEN PARTS I AND II—LITTLETON HOUSE GROUP

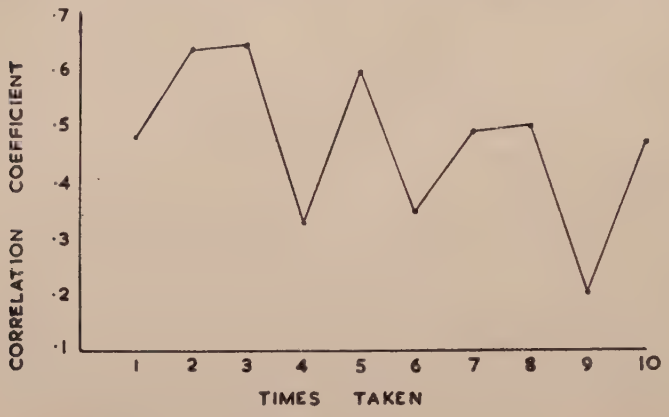
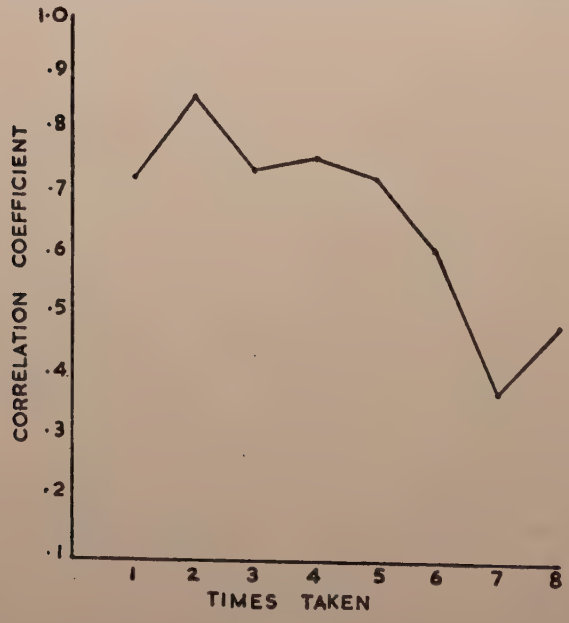


FIGURE 5 (*b*)

CORRELATION BETWEEN PARTS I AND II—W.E.A. GROUP



IV

DISCUSSION

If the mere taking of one test ten times has the effect of lifting a group of high-grade mental defectives well into the score range of a group of unselected subjects, it seems likely that with knowledge of results and with suitable explanation still more startling effects may be produced. It would be interesting to do further work along these lines, giving positive training, in addition to experience, in test-taking.

The present findings have immediate bearing on two practical problems: the weight to attach to psychological test scores when allocating to schools normal children who may or may not have been previously coached for such tests and the extent to which mental defectives can learn from experience and perhaps compete successfully, in certain fields, with more fortunate people.

It is evident that practice effects play as large a part in intelligence test taking as in many other skills, in spite of the fact that these tests are designed to assess *potential* ability—and do so fairly successfully when given to a group for the first time. The results found in this enquiry strongly suggest that the flattening of the "learning curve" observed with the W.E.A. students was due to the ceiling artificially imposed: it is probable that a test which offered the latter more scope for improvement would produce a rise in score comparable to that gained by the mental defectives.

The consistency of the group as assessed by comparing each testing with every other and with the pool of the scores, confirms the results found on this issue with the W.E.A. group. This has no bearing, of course, on the *validity* of the test, nor does it affect the unfortunate situation in which some members of a test group are more test-sophisticated than others. It is necessary for the satisfactory practical application of psychological tests though it is obviously very far from being sufficient.

The ease of Part II as compared with Part I is, in the writers' experience, largely a function of the low intelligence of the group chosen for the present enquiry. It is true that the mean of Part II is, in general, higher than the mean of Part I, but there is some evidence that this discrepancy tends to increase as the intelligence of the group decreases. Children gain markedly higher scores on the perceptual than on the verbal and numerical Part. It is not clear, however, whether this is related to their possession of greater powers of visual imagery than adults, or their lesser intelligence.

With unselected subjects there is a correlation of the order of 0·8 between the two Parts, some subjects doing better relatively on Part I and others on Part II.

The inability of the mental defectives to cope with Part I of AH 4 when contrasted with their facility on Part II and on the Matrix Test, emphasizes the importance of the medium used when assessing the mental ability of low-grade people—it being assumed that the psychologist aims to discover "how good" the subject is rather than "how bad" he is. At higher levels this finding suggests—as other findings have done—that an intelligence test which confines itself to one bias only may yield highly misleading results.

The upward trend of the progress curves suggests that in any case subjects with low intelligence may take a great many trials to reach the level beyond which they fail to improve. It was a pity that the end of term brought testing to a close with the eleventh session. On the existing data it is difficult to determine whether this slow but continuing progress is a property of low-grade groups on an intelligence test or whether the same result would be found with a brighter group on a more difficult test. The same doubt attached to the question whether more intelligent subjects will always improve to a greater extent than the less intelligent when the test is sufficiently difficult to allow full scope to the people originally gaining the higher scores.

Light may be thrown on some of these questions by repeating the experiment with groups of varying levels of intelligence and with a more difficult intelligence test. It might be of interest, too, to repeat the experiment with psychological tests other than those of intelligence.

V

Summary

(*a*) Twelve mentally-defective schoolboys between 14 and 16 years of age took test AH 4 once a week, ten times. They took also Progressive Matrices once, after they had finished the AH 4 testing.

(*b*) In addition to the above test scores, the writers had access to the boys' Binet I.Q.'s gained some years previously, and personal assessments from the Headmaster. Examination of these gave inconclusive results and they have, therefore, been omitted from this paper.

(*c*) The mean progress curves of the subjects showed (i) a tendency for improvement to continue throughout the ten testings, and (ii) marked differences between Part I (verbal and numerical) and Part II (diagrammatic) of the test, the latter starting at a much higher level and showing greater improvement.

(*d*) Inter-correlations between the testings were highly significant, surprisingly so in view of the selected nature of the group and the erratic nature of the progress of individual subjects. The consistency dropped strikingly, however, when the eight subjects with the lowest scores were treated as a separate group.

(*e*) The scores of the group on Progressive Matrices were startlingly higher than they would have been expected to be, had the boys taken this test *before* having taken AH 4 repeatedly, and without having completed examples prior to the test proper.

(*f*) The main differences between the test results of the group of mentally-defective schoolboys and the group of intelligent adults were as follows: (i) The schoolboy's individual progress curves were erratic and apparently unpredictable, whereas the W.E.A. students produced smooth curves with a well-marked trend. (ii) The low-grade group found Part II very much easier than Part I, judging from the number of questions attempted, the ratio between number attempted and number correct, and the number of questions omitted. The W.E.A. group had tended to maintain a higher ratio of correct to attempted on Part I than on Part II, throughout the testings, and these subjects had not omitted questions in either Part. (iii) The individuals in the duller group were far less consistent, both in their errors and in their correct answers, than the individuals in the superior group.

(*g*) It was suggested that the experiment be repeated with groups differentially pre-selected with respect to intelligence and with tests of different levels and media.

VI

References

1. Heim, A. W., and Wallace, J. G. (1949). The effects of repeatedly retesting the same group on the same intelligence test: I. Normal adults. *Quart. J. Exp. Psychol.*, **1**, 4.
2. Wallace, J. G. (1949). Results of a high-grade intelligence test applied to a university population. *Unpublished.*

The writers wish to express their warmest thanks to the Headmaster of Littleton House School for granting them facilities for carrying out the work, and for his kindness and help throughout the investigation. They are grateful, too, to Mr. E. G. Chambers for his assistance on the statistical side, and to Miss V. Horn for her help with the figures. They are especially indebted to Miss V. R. Cane and Dr. and Mrs. A. D. Harris for their invaluable help in the interpretation of results.

(*Manuscript received 3rd July*, 1949)

THE THRESHOLD FOR SUDDEN CHANGES IN THE VELOCITY OF A SEEN OBJECT

BY

W. E. HICK

(*From the Medical Research Council Applied Psychology Unit, Cambridge*)

The difference threshold for the velocity of a seen object was measured by the method of constant stimuli, using two categories. An approximate correspondence with Weber's law was found, the divergence from it appearing, in general, as an increase of the threshold at both ends of the range of initial velocities. The Mean Threshold (0·5 probability of perception, corrected for guessing) was, in favourable conditions, about 12 per cent. of the initial velocity. Whether the stimulus was an increase or a decrease of velocity made no marked difference. With two moving objects, which converged, crossed, and then diverged, both suffering the same change of velocity, the threshold was higher.

Velocity changes as low as 2·5 per cent. elicited a significant proportion of correct responses. Some theoretical points in connection with this are discussed. Responses to blank stimuli showed a strong tendency to guess "slower," which tendency differed significantly in degree between most of the experimental conditions. Tests with reduced exposure times showed that exposures could be as short as 0·5 second (the velocity change occurring in the middle of the exposure) without appreciable detriment.

I

Introduction

The estimation of rates of change of quantities is an essential part of many forms of prediction. In a number of important cases the quantities are perceived visually, and are of the nature of positions or shapes of objects. The particular task which suggested this study is that of tracking, in which, by the operation of a control handle and its connected mechanism, two objects are required to be kept in coincidence in spite of a continuous tendency to diverge. Other tasks, similar in principle, and on which the present findings may have a special bearing, are driving in convoy at night, and flying in formation. Many more examples could be given, the essential common feature for this purpose being that the primary stimulus is a spatial dimension and that efficient control demands anticipation of its future value on the basis of its present rates of change—in other words, the ability to judge velocity and, if possible, higher derivatives.

Little previous work seems to have been done on this subject. Dimmick and Karl (1930) measured the effect of exposure time on the least perceptible velocity, and other investigators have studied the influence of illumination, position in the visual field, background, etc., but no reference has been found dealing with the ability to discriminate between different successive velocities of a moving object. This is essentially the subject of the present study. It can be regarded as a measurement of the difference threshold for velocity, or as a preliminary approach to the measurement of the acceleration threshold.

II

Methods and Apparatus

Two methods, using quite different apparatus, were employed. In the first (called the "drum" method) an endless band of paper some 6 ft. long passed over a drum, in front of which was a mask having a narrow slit. The moving object was provided by a line on the paper, a short segment of the line being visible through the slit. The motion of the object was horizontal, and consisted of periods of between 2 and 4 seconds of constant velocity, the changes of velocity being instantaneous,

The subject was seated with his eyes approximately 18 in. from the mask. He was provided with a knob, to be moved to right or left according to the direction of the momentary acceleration. The instruction was not to hold the knob over, but merely to apply a brief pressure in the correct direction as soon as a change of velocity was perceived.

In the first trials the object was liable to accelerate or decelerate when moving either to the left or the right, and a deceleration would occasionally be sufficient to cause a reversal of motion. In order to explain the task to the subject he was asked to imagine that the object, when it changed speed, had been given a push, and he was to press the knob in the same direction as the imaginary "push." Although a few subjects grasped the requirement at once, and most did so after some practice and demonstration, it was soon evident that the unfamiliarity of the concept of positive and negative acceleration was causing errors that had nothing to do with perception. The task was therefore simplified by making the spot move only from right to left, with no reversals. As soon as it had reached the left-hand end of the slit it began again at the right-hand end.

Even so, this method exhibited certain defects in practice, the chief of which was that, in order to get a measure of the reaction time, it was necessary that the stimuli should appear at irregular intervals. The result of this and other factors was that stimuli many times greater than threshold value were frequently missed, which led to complications in assessing the results. The technique of "forced guessing" could not be used, and therefore there was no control over the degree of confidence demanded by the subject before he would respond. Moreover, it was impracticable, with the drawing instruments at hand, to produce a standard set of stimuli in sufficient numbers to prevent the sequence on each band being liable to be learnt. The lines were, in fact, drawn haphazard, within general limits, and this prevented learning.

The second method (called the "oscilloscope" method) was devised to reduce these defects. To a standard Furzehill oscilloscope with D.C. amplifiers was coupled what was essentially a Miller circuit. By means of a multiway switch, condensers could be selected to give sweep speeds of the spot of 1·32, 3·12, 7·02, 11·4, 22·98, 49·5 and 89·3 mm. per second on the screen. These were measured by photographing the spot on moving film, and the constancy of the speed across the tube face was checked by the same means.

The circuit was so arranged that extra resistors could be connected in parallel with that through which the condensers charged, and these, by changing the rate of charging, changed the speed of the spot. The velocity increments obtainable in this way were fixed percentages of the initial velocity, and ranged from 2·5 to 75 per cent. for increases and from 2·44 to 42·9 per cent. for decreases. These percentages were checked partly photographically and partly by stop-watch (using the lower initial speeds).

The switching in of the parallel resistance was performed automatically when the spot reached a predetermined position. A silent switch was provided, out of sight of the subject, for the experimenter to pre-select "faster" or "slower." This automatic switching device, which consisted essentially of a triode and two relays, was given a separate H.T. supply, to eliminate any possible interference which might have jerked the spot at the moment of operation.

In order to extend the range of the experiments two further elaborations were made. One was to produce the appearance of two spots by means of a commutator, driven by a motor at well above flicker speed, which repeatedly reversed the input to the oscilloscope from the Miller circuit. The two spots approached one another from opposite sides of the tube and crossed in the middle. The other was a pair of shutters which could be placed so as to cover all but the middle portion of the tube face, the aperture being adjustable as required. The purpose of this was to examine the effect of altering the exposure of the spot. Two difficulties were encountered here. The spot-position at which the automatic switch operated was subject to random variations from occasion to occasion, with a standard deviation of about 0·9 mm., not to mention a slow drift, and therefore it was not practicable to reduce the aperture to less than 1 cm. In addition, it was clearly essential to give the subject proper warning of when the spot would appear; this problem was solved by incorporating in one of the shutters a piece of transparent plastic, moulded so as to break up the light; through this a faint irregular glow from the approaching spot could be seen, but no clue as to its velocity could be gained.

The experimental procedure with this apparatus (the oscilloscope) was as follows. The subject was seated with his eyes at approximately 21 in. from the display, and told to say "faster" or "slower" according to his judgment of the change of speed. Reaction

time was not, of course, recorded. At the beginning of a series he was told approximately where on the tube the change of speed would occur, and that he would know by the click of the relays each time that it *had* occurred; conditions were therefore as favourable as possible for expectancy and attention.

The subject was also told, untruthfully, that there would always be *some* change. These blanks amounted to about 15 per cent. of all stimuli (except in the case of the reduced-aperture experiment, where they were not used at all), and, as will be seen later, they contributed an indispensable item of information.

The general procedure was to begin by giving the subject a few practice trials, and then to present the standard series of stimuli in descending, and then ascending, order of magnitude, at each value of initial velocity. In fact, this rule was not strictly adhered to, because it was thought that, since the method of forced guessing (no third category) was being used, a considerable number of not consciously perceptible stimuli in succession might have a bad effect on motivation; for this reason a relatively easy stimulus was always introduced after three or four "impossible" ones. The series of initial velocities was only given once to each subject, at least with the same conditions, but approximately equal numbers received it in ascending and descending orders.

Six subjects were used in the drum method and eighteen others with the oscilloscope, but not all the eighteen were tested in all the conditions. The conditions not already mentioned refer to the positions of the velocity change in the case of the double spot. These were:—

(1) Spots 2 cm. apart, converging.

(2) Spots 2 cm. apart, diverging.

(3) Spots just crossing each other.

III

Results

The definition of the threshold which was chosen was that of the so-called Mean Threshold. It is that value of the stimulus which makes the number of smaller stimuli perceived equal to the number of greater stimuli not perceived. It can be given a physical meaning by assuming that there is, at any instant, a definite and absolute threshold, but that it varies from instant to instant; then the Mean Threshold is the *median* value of the instantaneous threshold. Consequently it is an estimate of the stimulus value which has a 0·5 probability of being perceived, notwithstanding that this stimulus may, in the experiment, have been perceived on much more or less than 50 per cent. of occasions. Such a discrepancy has then to be ascribed to sampling error.

This method has the advantage of making some use of all the data, and of reducing the labour of computation to the realm of practicality; indeed, where the stimuli are separated by equal intervals and are presented equal numbers of times, the work amounts to little more than counting the total number perceived. Unfortunately these conditions could not be satisfied in the present case.

On the other hand, it is essential to the method that the stimuli should cover the whole transition region; that is, they must range from the "always seen" to the "never seen." The data in the present study required a certain amount of manipulation to bring them into reasonable conformity with this requirement. In the case of the drum method, even the largest velocity changes used (up to about thirty times the threshold value) were occasionally missed. In fact, above the apparent transition region the probability of seeing a stimulus could be seen to have little, if any, relation to its magnitude. Doubtless the reason was something like inattention or confusion. These probabilities were estimated (they were of the order of 0·9), and it was assumed that the probability of actually responding to a given stimulus was the product of the probability of its being above the instantaneous threshold and

that of the subject being in a suitable state to respond to any stimulus. Thence a correction to the observed probabilities over the whole range of stimuli (for each initial velocity) was obtained, having the effect of increasing them by some 10 per cent. (except, of course, those that were already unity).

In the oscilloscope experiments it was necessary to correct the observed frequencies for chance guessing. The responses to the blank stimuli were used for this purpose; they showed a strong over-all preference for "slower" rather than "faster," to the extent of nearly 75 per cent. in one particular condition. Had the conventional assumption of an equal chance of guessing one or the other been used the results would have been grossly distorted.

With the two-category technique it is, of course, impossible to give a stimulus so small that it can be said with certainty to be never seen. The correction for chance guessing can, at best, only be right on the average. Taking all the oscilloscope data for the smallest velocity change (2·5 per cent.), an appreciable but not significant preponderance of correct reports was found. There were several individual cases of highly significant frequencies of correct responses to this small stimulus, but that is natural in any large collection of data, and need not indicate any real effect. However, the unavoidable result was that the corrected frequency of perception of the smallest stimulus was sometimes positive and sometimes negative, but seldom zero; hence the conditions for computing the Mean Threshold by the method chosen were not exactly satisfied. The thresholds, as finally calculated, are therefore probably a little too large, but the error is certainly very small.

Theoretically, it is possible to estimate the standard error of the Mean Threshold, since the variable term in the formula is essentially the sum of a number of binomially-distributed quantities. But it appeared that there were difficulties in applying the method usefully to the present case, and it was thought that, as a number of thresholds were calculated, the variability would be sufficiently evident for practical purposes.

The Mean Thresholds obtained with the "drum" method are given below (Table I), in percentages of the initial velocities. The latter are in degrees per second at the eye, and are group means.

TABLE I

Initial velocity	0·11	0·38	0·77	1·28	2·31	4·18
Mean Threshold (per cent.)	133	41	46	31	21	12

The only interest lying in these figures concerns their marked disagreement with those obtained by the oscilloscope method, which are given below (Table II). It can be seen that the tendency to miss perceptible stimuli is very great at the lowest speed, but declines rapidly as the speed is increased, until, in the region of 4 degrees per second, it is inappreciable. The reason is probably that, at the higher speeds, the spot made a succession of fairly rapid sweeps across the display, and the subject properly inferred that a velocity change would occur somewhere in each sweep. Consequently he had a clear warning signal, and, in addition, the circumstances would favour an attitude approaching that of forced guessing. At the low speeds, on the other hand, the spot drifted slowly across, suffering a number of velocity changes in its course, and the subject might be both less attentive at the relevant instants and more inclined to wait for an unmistakable stimulus before responding. This was one reason for repeating the experiment with the oscilloscope apparatus.

With the drum method there were 29 responses in the wrong direction out of a total of 962 (3 per cent.). A significant majority of these wrong movements were in response to the second of a pair of stimuli, both of the same direction. Since there was only one case of two wrong responses in the same direction, it is likely that the principal cause of wrong responses here was a tendency not to do the same thing twice in succession, as found by Thorndike (1927) and Telford (1931) among others.

TABLE II

MEAN THRESHOLDS AS PERCENTAGES OF INITIAL VELOCITIES, FROM OSCILLOSCOPE DATA

(Degrees per second at Eye)

(*a*) *Single Spot: Velocity Increment.*							
Initial velocity ..	0·15	0·35	0·81	1·31	2·64	5·69	10·25
Mean Threshold ..	12·6	9·9	9·0	9·6	9·6	16·0	13·2
(*b*) *Single Spot : Velocity Decrement.*							
Mean Threshold ..	15·0	14·7	14·4	13·5	9·3	12·2	20·0
(*c*) *Double Spot (Converging) : Increment.*							
Mean Threshold ..	17·5	15·2	11·4	8·3	7·2	11·7	20·8
(*d*) *Double Spot (Converging) : Decrement.*							
Mean Threshold ..	14·6	13·8	17·9	18·1	19·9	19·3	19·1
(*e*) *Double Spot (Diverging) : Increment.*							
Mean Threshold ..	22·1	—	21·2	—	14·9	—	19·3
(*f*) *Double Spot (Diverging) : Decrement.*							
Mean Threshold ..	12·7	—	13·4	—	11·4	—	12·9
(*g*) *Double Spot (Crossing) : Increment.*							
Mean Threshold ..	17·1	—	13·3	—	13·7	—	10·7
(*h*) *Double Spot (Crossing) : Decrement.*							
Mean Threshold ..	15·9	—	13·6	—	10·6	—	10·9

Table II shows that the Mean Threshold obeys Weber's law approximately, the principal divergence from the law appearing, in most of the conditions tried, as an increase of the threshold at both ends of the range of initial speeds. This very common finding is clearly inevitable in the present case if the range is extensive enough. Very slow motion, as of the hand of a clock, ceases to be appreciated as such, and the only perceptible change of velocity is one which raises the velocity to the level of being detectable as velocity, and this could be a very large relative increase. Decreases are only perceptible in terms of remembered displacements and time intervals, one or both being necessarily liable to large error. Very fast motion likewise ceases to be appreciable as motion. It is true that, in these experiments, the extremes were not reached, but it appears that they were approached. At the lowest speed the spot did appear motionless at a casual glance; while, at the highest speed, the limitation may, as discussed below, have been the short time during which the spot was in view.

The results from the single-spot experiment indicate that the threshold is in the region of 12 per cent. of the initial velocity; this is confirmed by the appropriate portion of the data from the subsequent reduced-aperture tests, and may, therefore, be taken as reasonably reliable. With the double spot, however, the results in the

various conditions are somewhat divergent, and it is difficult to know how to interpret them. There is some general tendency for the thresholds to be higher than with the single spot; in this connection it is interesting to note that most subjects concentrated on one spot, the other being regarded as more or less confusing and distracting. To have looked at both would have been, in effect, to employ parafoveal vision, which is usually said to be more sensitive to movement than foveal vision; but the latter may be found to be a better detector of change of velocity—in other words, a better measurer of a velocity that is well above threshold in absolute magnitude. The present results only raise this question, of course; they do not answer it.

A clearer picture of the effects of the different conditions is obtained when the figures for increments and decrements are pooled, as in Table III.

TABLE III

	Initial velocity ..	0·15	0·35	0·81	1·31	2·64	5·69	10·25
(*a*)	*Single Spot.* Mean Threshold ..	13·8	12·3	11·7	11·6	9·5	14·1	16·6
(*b*)	*Double Spot—converging.* Mean Threshold ..	16·1	14·5	14·7	13·2	13·6	15·5	20·0
(*c*)	*Double Spot—diverging.* Mean Threshold ..	17·4	—	17·3	—	13·2	—	16·1
(*d*)	*Double Spot—crossing point.* Mean Threshold ..	16·5	—	13·5	—	12·2	—	10·8

The probabilities of guessing "faster," as estimated from the responses to the blank stimuli, were, for the various conditions:—

(*a*) *and* (*b*)	(*c*) *and* (*d*)	(*e*) *and* (*f*)	(*g*) *and* (*h*)
0·32	0·45	0·25	0·46

Of these, 0·45 and 0·25 differ significantly at the 0·5 per cent. level, and the other possible pairs (except 0·45 and 0·46) also differ significantly, though less so. Since the same subjects were used in conditions (*a*) to (*d*), and some of them performed in the other conditions as well, the differences cannot be attributed to personal idiosyncrasies. The cause must presumably have to do with the different conditions, but what it is can only be conjectured. Naturally, when a subject is asked to arrive at a decision, all rational grounds for it having been carefully removed, he must base it on irrational grounds—on such symbolic suggestions as he may be able to glean from the total situation. It is always of interest when such suggestions are found to operate in the same way on the majority of subjects. It may be noted that 289 of the total of 805 blanks were guessed as "faster," which is so much less than the pure chance expectation as to be far beyond the range of the usual table of χ^2. Yet the 2·5 per cent. *increase* of velocity was correctly reported 98 times out of 184, which indicates, with a significance better than 0·1 per cent., that even this small change is capable of reaching the cerebral apparatus.

IV

Discussion

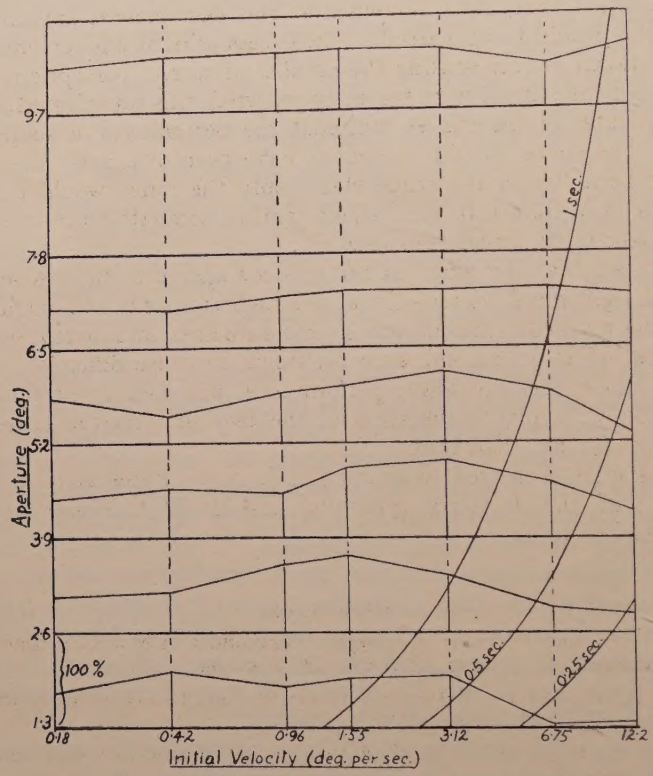
These results illustrate the advantages and disadvantages of the two-category or forced-guessing method. It eliminates the very uncertain factor of the degree of confidence the subject requires in his judgments before he will admit that they are "conscious." It would be better if we could control, or at least measure, this factor. But eliminating it has the justification that many of the judgments made in the course of a skilled activity are not conscious; in fact, the circumstances must often amount to a requirement for forced-guessing. The over-all performance must be considerably affected by judgments which, if they were made in isolation in experimental conditions, the individual would have no subjective confidence in, but which, on the whole, are more often right than wrong.

Then, behind this, there is the prevalent notion that there is a real physiological threshold, towards which observed thresholds tend with long practice. Many elementary physiological experiments and demonstrations have contributed to this notion, which implies that the nervous system is a true ankylotic system, with something approaching an absolute discontinuity between adequate and inadequate stimuli. On the other hand, experiments of the kind described here suggest that, if there is a definite physiological threshold, it is sometimes very far below what would normally be called the psychological threshold. It is possible that even the smallest stimulus, provided it has a real physical existence—i.e. contains at least one quantum of energy—is capable of producing a bias towards making the correct response. The bias may be extremely small, and yet produce a significant effect in a long enough series of trials. It has been argued by de Vries (1948) that sensory cells and nerve fibres are as sensitive as it is possible for them to be without suffering an undue number of spontaneous excitations. If this is so—or, indeed, if it be accepted that spontaneous excitations, due to random processes at the molecular level, do occur—it follows that a stimulus too small to discharge a cell when it is in its average condition will, nevertheless, slightly increase the probability of excitation. It is evident, of course, that even if a sensory cell succeeds in discharging, the resulting signal has to run the gauntlet of many other randomizing processes before it eventually finds expression in activity.

It remains an open question why a large or relatively unmistakable stimulus gives rise to a conscious sensation of having responded correctly, while a very small stimulus, even when it evokes the correct response, leaves the subject feeling that he has made a pure guess. A reasonable hypothesis seems to be that the deciding factor is immediate memory; if the subject's memory image of the stimulus suggests to him the same response as the one he actually made, he may fairly conclude that he was right; but a very small stimulus may produce a trace which is obliterated almost at once, leaving the subject unable to decide whether he really perceived it or not. The alternative possibility that the feeling of confidence does not have to await this review of the memory image, but is born with the response, and that the image is only a kind of rationalization projected into the past, is probably only an apparent alternative—another way of saying the same thing. The fact that memory images in general can be searched for details not originally asked for does not help, because it can always be argued that these details caused changes of attitude or "set" which, though not visible, were just as much entitled to be called responses as the overt movement. Although the writer believes that this kind of argument can be fitted into a consistent system, the cornerstone of which is the thesis that knowledge is fundamentally operational in character, it is unnecessary to pursue

the matter here; in any case, the question raised is possibly not one that experimental psychology, as such, can answer. Even the fact that, in the present experiments, "easy" stimuli were usually responded to in a loud and cheerful voice, contrasting with the dull monotone evoked by the more difficult ones, means little more than that different stimuli produced different responses.

FIGURE 1



With regard to the reaction time records, the principal finding was the usual one that the more difficult the stimulus the longer was the reaction time. As explained above, the reaction times were calculated from the "drum" data, only those for the correct responses being taken account of. On the whole, the times were longer for the initial velocities, which accords with the higher thresholds observed in these conditions. Mean reaction times varied from 2·1 seconds to 0·38 seconds. The actual figures do not appear to convey any important information, and are therefore omitted.

The essence of the results obtained in the reduced-aperture experiment is given in the diagram. The velocity change was set at approximately 12 per cent., this being about the value which would result in something like 50 per cent. of correct perceptions—i.e. it was near the Mean Threshold—and would therefore be expected to yield the most information. Eight subjects were used, and the technique was that of forced-guessing, as before. Starting with the largest aperture the subjects were given equal number of increments and decrements at each initial velocity, starting with the lowest; in other words, the initial velocities were presented in

ascending order, and repeated at each aperture, the latter being in descending order. Theoretically, a fully randomized order would have been better, but there were several practical objections. In the first place all these tests were monotonous and fatiguing to the eyes, and the forced-guessing procedure is an unsatisfying one to the subject; from preliminary trials it seemed that a regular progress from the easier to the more difficult was the least disturbing to the subjects, and the most likely to secure their co-operation.

It was assumed that, since incremental and decremental stimuli were pooled, random guessing would be unbiassed. The excess of right answers over 50 per cent. was therefore taken as representing the number of correct perceptions; for example, if 75 per cent. of the stimuli were correctly reported, this was plotted as 50 per cent. Thus, the ordinates of the curves represent the percentages of supposedly correct perceptions. The curves would, of course, have been of exactly the same shape if the data had been left in the crude state; only the range would have been from 50 to 100 instead of from 0 to 100. Each marked ordinate represents a percentage derived from about 60 actual responses.

It can be seen that the effect of the reduced aperture only becomes noticeable at the bottom right-hand corner—i.e. at the highest speeds and smallest apertures. The curved lines marked in seconds are iso-exposure lines, and there is some suggestion that exposures below 0·5 second were beginning to cause difficulty. It is evident that aperture alone does not affect performance; nor does speed alone. Exposure is the most likely simple combination of the two, and on the present data it is impossible to go further than that.

In conclusion, it is desired to thank Dr. J. A. V. Bates and Dr. A. Carpenter for assistance with the electronic apparatus, and Mr. D. C. Fraser for carrying out some of the earlier tests.

V

References

1. Dimmick, F. C., and Karl, J. C. (1930). The effect of exposure time upon the RL of visible motion. *J. Exp. Psychol.*, **13**, 365–369.

2. Telford, C. W. (1931). Refractory phase of voluntary and associative responses. *J. Exp. Psychol.*, **14**, 1–36.

3. Thorndike, E. L. (1927). The refractory period of associative processes. *Psychol. Rev.*, **34**, 234–236.

4. de Vries, H. (1948). Minimum perceptible energy and Brownian motion in sensory processes. *Nature*, **161**, 4080, 63.

(*Manuscript received 28th September*, 1948.)

PROCEEDINGS OF THE EXPERIMENTAL PSYCHOLOGY GROUP 1948-9

April 24–25, 1948. Extended Meeting at Cambridge. *1st Session :* Symposium on Artistic Expression in Relation to Mental Disorders, by E. B. Davies (by invitation), G. Westby (by invitation) and R. W. Pickford. *2nd Session :* "Parallels between the Conditioned Reflex and the Phi-Phenomenon," by George Humphrey (by invitation). *3rd Session :* "Experiments on Repeated Testing of Intelligence," by A. W. Heim. "Preliminary Studies of the Changes in Skilled Performance associated with Increasing Age," by A. T. Welford.

July 24, 1948. Social Meeting in Edinburgh for Members and Guests attending the 12th International Congress of Psychology.

December 4, 1948. (London): "Some Experiments on the Visual Perception of Size," by R. B. Joynson. "Non-Verbal Reasoning in Cases of Cerebral Lesion," by A. E. Sladden.

January 8, 1949. 2nd Annual General Meeting in London. "The Functional Significance of Electrical Systems in the Brain," by W. Grey Walter (by invitation).

March 30–31, 1949. Extended Meeting at Oxford. *1st Session :* "Form Perception in Dark Adapted Vision," by B. Semeonoff. *2nd Session :* Symposium on Prefrontal Leucotomy and Related Problems, by C. W. M. Whitty, P. Glees and MacDonald Tow (all by invitation). *3rd Session :* "Experimental Sensory Dissociation," by D. C. Sinclair (by invitation). "Apparent Fluctuations in a Sensory Threshold," by R. C. Oldfield.

July 23, 1949. Meeting at Cambridge in association with Symposium on Animal Behaviour arranged by the Society for Experimental Biology and the Institute for the Study of Animal Behaviour. *1st Session :* "Researches at the Yerkes Laboratory of Primate Biology," by K. S. Lashley (by invitation). "Perception and Insight," by K. S. Lorenz (by invitation). *2nd Session :* Short *interim* reports of work in progress at the Cambridge Psychological Laboratory, by Nancy Harris, P. H. R. James, C. Poulton and C. B. Gibbs (all by invitation).

September 15–16, 1949. (Oxford): *1st Session :* "Work in Progress at the Psychological Department of the National Hospital, Queen Square: An Informal Review," by J. McFie (by invitation), M. F. Piercy (by invitation) and O. L. Zangwill. *2nd Session :* "Cybernetics: How Nervous Structures have Ideas," by W. S. McCulloch (by invitation), followed by Discussion. *3rd Session :* "Degrees of Assurance," by B. Babington Smith.